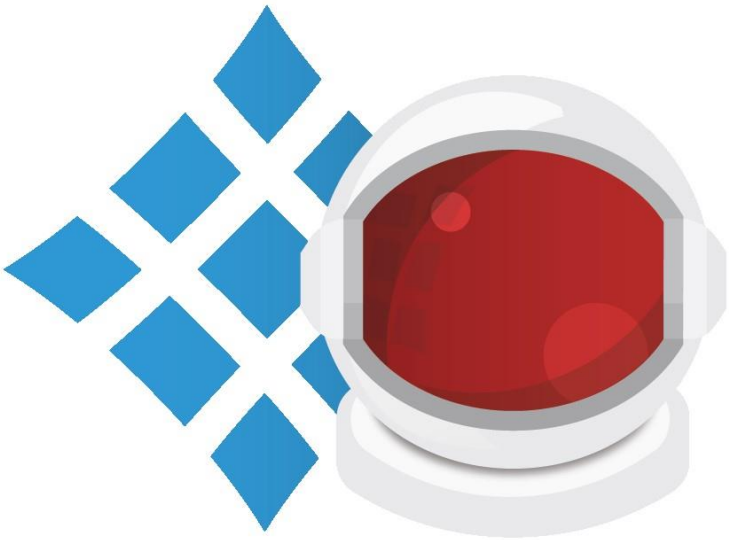


Spectronaut®

User Manual

1 General Information

1.1 Scope of Spectronaut® Software

Spectronaut® is a commercial software package aimed at analyzing data independent acquisition (DIA) proteomics experiments. Spectronaut can quantitatively profile hundreds to several thousands of proteins in one experiment. Large experiments with several conditions and replicates consisting of up to tens of thousands of LC-MS runs can be analyzed.

Spectronaut can analyze DIA data without the use of a retention time calibration kit. However, the addition of the iRT Kit is highly recommended as it ensures calibration on difficult matrices and allows for detailed quality control readouts.

1.2 Spectronaut® 18.0 Key Release Features

- Greatly improved quantification (12% more proteins with CV<10%)
- Linux support via command line (direct loading of Bruker & Thermo files, Sciex via HTRMS)
- Improved protein identifications (5% more on average)
- Added search archive support for directDIA+
- Greatly expanded command line capabilities
- Streamlined & improved licensing for cloud installations.
- Added MaxLFQ support for SNE-Combine
- Violin plot option for all Box Plots
- Improved error and warning feedback during pipeline processing
- Added “Quantified” summary node in Analysis Overview
- Directly use search archives (.psar) in regular DIA analysis
- Add identifications per cycle overview plot on run level.
- Changed all files exported as .XLS to .TSV
- New default quantification setting (via DIA analysis → Quantification → Quantification window)

1.3 Computer System Requirements

Spectronaut® is only available for Windows operating systems. Command line operation is also supported (see Section 0). The minimum and recommended system specifications are described in Table 1.

Table 1. Operating system specifications

Specifications	Minimum	Recommended *
Operating System	Windows 10, x64 / Ubuntu 22.04 LTS	Windows 10 or higher, x64 / Ubuntu 22.04 LTS
CPU	Intel® Core™ CPU, 2.7 GHz (4 cores) or similar	Intel® Core™ i7 4770, 3.4 GHz (octa core) or similar Intel or AMD CPU with 4 or more cores
Hard drive	200 GB free space	2 TB of hard drive space or more (2x data set size). The required temporary storage space scales with the library size and the number of run files being analyzed.
RAM	16 GB	64 GB or more (16 GB of memory or more (1 precursor in 1 run amounts to ~0.5 KB of RAM)
Software	.NET 6.0	.NET 6.0 or higher / .NET 6.0 for Linux

**For a typical experiment of 200 runs with a 100k precursor library*

The memory growth for a given experiment can be estimated using the following equation:

$$RAM_{GB} = 5.0 + \frac{0.6 * n * r}{1024^2}$$

where n is the number of precursors in the library and r is the number of runs in the experiment. The baseline memory consumption (estimated at 5 GB in this equation) can vary by vendor and gradient length. To get an estimate for the baseline, you can analyze a single, representative raw file with the target library. Spectronaut has been successfully tested running 1,000 2 h DIA raw files with a library of 200,000 precursors on a 128 GB RAM system.

1.4 Post Installation Recommendations

Performance improvements after the installation:

1. **Directories:** Spectronaut® will set all directories in the *C:* drive by default. However, it is likely that the *C:* drive has a limited storage capacity. Thus, we **strongly recommend** changing the Temporary Directory and the Local Search Archives directory to a local destination with enough free memory. To do that, go to the Settings Perspective → Global → Directories (Figure 1).

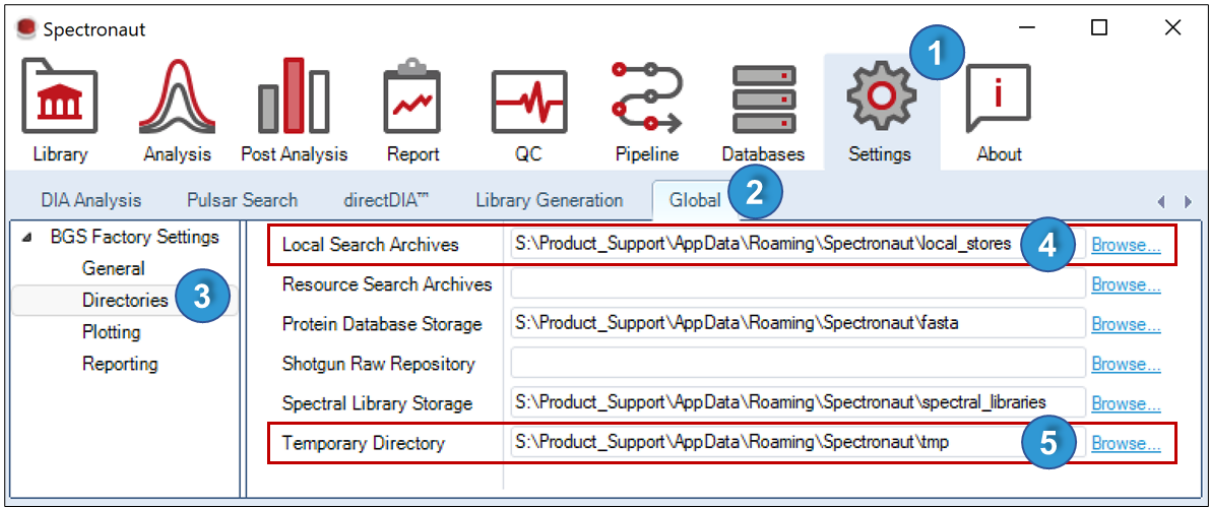


Figure 1. Change the default location of Local Search Archives and Temporary Directories to local destinations with enough storage capacity.

2. **Network drives and virtual machines:** We strongly recommend running Spectronaut locally, i.e., having the resources (especially run files, Search Archives, and temporary directories) on a local drive. The use of virtual machines is feasible but not advised. A failure in the connection to any network drive can cause Spectronaut to abort the process due to third-party library dependencies.

3. **CPUs:** Spectronaut is designed to perform highly resource-intensive tasks, especially when running searches with Pulsar, the Biognosys' proprietary search engine. For resource management purposes, you can set a maximum number of CPUs for Spectronaut to use. To do that, go to the Settings Perspective → Global → General → CPU Affinity, and uncheck one or more of the CPUs (Figure 2). Be aware this may prolong the analysis time.

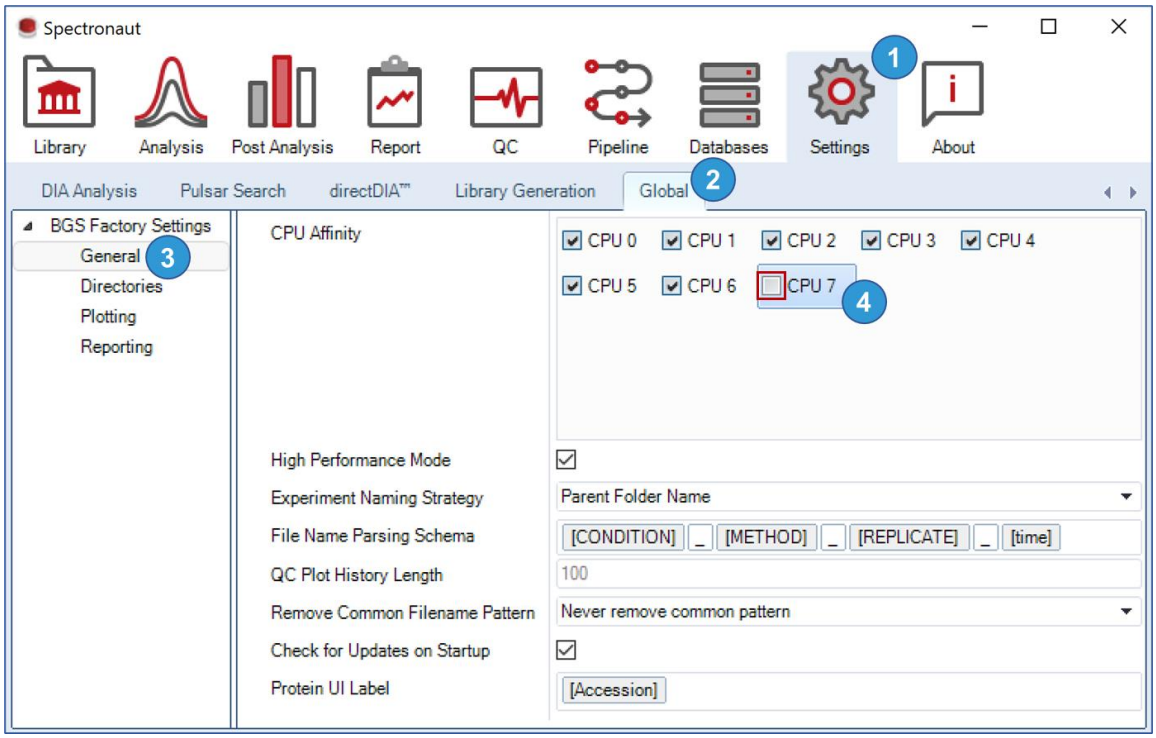


Figure 2. Designate the CPUs that Spectronaut can use.

1.5 Spectronaut on Linux OS

Starting from version 18, Spectronaut installer for Linux OS is available. Spectronaut Linux version is run via command line. Spectronaut experimental files (.sne fies) are exchangeable between operating system, which allows to get all Spectronaut GUI visualizations in the Spectronaut Viewer on Windows OS. Below you can find further specification of Linux version of Spectronaut:

- Support for library generation, HTRMS conversion, directDIA workflow, library-based DIA analysis and SNE combine workflow
- Experimental settings schemas are supported by a JSON file or schema file in the input folder.
- Global settings edition is currently not supported by JSON file, only available as settings schema.
- For the library generation, Mascot search results are not supported
- Support for the following post analysis reports: Experiment Overview, CVs below X, Run Identifications

Prerequisites for the installation of Spectronaut on the Linux OS:

- .NET 6 – for the installation instructions, refer to the appropriate section of .NET documentation: https://learn.microsoft.com/en-us/dotnet/core/install/linux-ubuntu
- The user running Spectronaut has to have a home directory (/home/<username>) with read and write permissions to it.

For the installation run the following command:

sudo dpkg -i spectronaut-VER.deb (where VER indicates version of Spectronaut being installed).

For running Spectronaut run the following command:

spectronaut <options> (where <options> are appropriate options for running Spectronaut, which can be checked in a Windows batch file. Output files should be by default placed in user's working directory ".").

1.6 Supported Mass Spectrometers

Spectronaut® supports mass spectrometry DIA data from Thermo Scientific™, SCIEX, Bruker, and Waters. The specific supported models are:

- Thermo Scientific™ Orbitrap Exploris™ Series

- Thermo Scientific™ Q Exactive™ Series
- Thermo Scientific™ Orbitrap Fusion™ Series
- SCIEX TripleTOF® Series
- Bruker impact II™, timsTOF and timsTOF Pro
- Waters Xevo® G2-XS QToF, Synapt G2-Si

Linux Spectronaut installation supports analysis from Bruker and Thermo vendors, instruments as listed above.

1.7 Supported Data Acquisition Methods

Spectronaut® analyzes a large variety of DIA data. Minimum requirements are a reversed phase chromatography with either a linear or nonlinear gradient that spans at least 10-35% acetonitrile. Methods acquiring MS1 and MS2 scans are recommended; methods acquiring only MS2 scans are also supported. For more information on setting up a DIA acquisition method, please refer to Box 1. The cycle time of the DIA method should be in the range of 2-3 seconds depending on the peak width of the chromatography used. MS1 as well as MS2 scan ranges can be segmented. The MS2 scans should cover at least 500-900 m/z of precursor range. More specifically Spectronaut supports HRM™ (Bruderer et al., 2015), WiSIM-DIA (Kiyonami, 2014), AIF (Geiger et al., 2010), SWATH™ (Gillet et al., 2012), SWATH™ 2.0 (Lambert et al., 2013), SONAR™ (Moseley et al., 2018), BoxCar (Meier et al., 2018) DIA, FAIMS Pro (Box 2; (Bekker-Jensen, Martinez-Val, et al., 2020)) dia-PASEF (Box 3), and HDMS^E (Distler et al., 2014). RTwinDIA (Li et al., 2019) is supported as well as the staggered/shifted window method (Amodei et al., 2019) upon MS2 demultiplexing via the HTRMS converter (co-installed with Spectronaut; see Section 4). Although fractionation in DIA experiments is not recommended (Box 9), it is supported, including gas phase fractionation. Multiplexed DIA (Egertson et al., 2013) and MS^E (Silva et al., 2006) are not supported. In case you experience technical problems with the software, or if you have feature suggestions, please contact us via Help Center.

Box 1. Recommendations for DIA Method Setup

Although MS1 information in DIA is not strictly required, it is highly beneficial. MS1 information boosts the sensitivity and is a very useful orthogonal information that significantly improves peak picking and scoring, leading to a higher number of identifications (roughly 20-30%). It also hardly adds any time to the cycle time.

In a DIA analysis, the MS1 XIC is reconstructed and evaluated by a set of different scores that focus mostly on mass accuracy, isotopic pattern, XIC shape, and intensity. Together with the MS2 level scores, they

are weighted and incorporated into the final peak scoring. In a peptide-centric approach, the MS1 information will add to the scoring, but the scoring is not dependent on an MS1 signal being of good quality or even present at all. For example, the dynamic range of certain biological samples can cause some peptides to remain undetected in the MS1 scan. In such cases, if the MS2 signal is good enough, Spectronaut can recover that peptide ion information.

Differently, a spectrum-centric approach strongly relies on the good quality of both MS1 and MS2 signals for precursors identification. By default, the MS1 signal is not used for quantification, unless actively chosen in the DIA analysis settings.

If you are interested in further insight, please learn more in our MCP article on the optimization of experimental parameters in DIA (Bruderer et al., 2017).

Box 2 Ion Mobility DIA with Thermo Scientific™ FAIMS Pro™

With the introduction of Spectronaut 14 we greatly increased the support for the FAIMS Pro™ device for ion mobility (IM) filtering. DIA methods with FAIMS usually come in two main themes. Single CV methods and multi CV methods. While single CV methods can be used with any already existing spectral library (including those not recorded with a FAIMS device). DIA methods with multiple FAIMS windows require matching IM (CV) annotation to be present in the spectral library to function.

To generate spectral libraries for FAIMS DIA we recommend the usage of Pulsar, our integrated search engine. Pulsar supports the search of both FAIMS DIA and DDA methods for library generation.

Alternatively, one can also use Proteome Discoverer for generating spectral libraries from FAIMS DDA.

	Single CV	Multi CV
Standard Library	✓	✘
FAIMS Library	✓	✓
directDIA	✓	✓

To analyze DIA data with multiple CV windows, one requires a library with CV annotation per precursor (ideally measured on a DDA or DIA method with similar CV settings). The analysis of single CV DIA files does not require any specialized libraries and can therefore be done with pre-existing libraries that were generated on DDA or DIA data without the use of a FAIMS Pro™ device.

Alternatively, you can also analyze FAIMS DIA runs without the need for a spectral library using directDIA.

Box 3. Ion Mobility dia-PASEF with Bruker timsTOF Pro™

Spectronaut supports dia-PASEF workflows. Spectronaut processes dia-PASEF data based on high-precision ion mobility (IM) calibration workflow which is conceptually similar to retention time calibration (Escher et al., 2012).

To analyze dia-PASEF data, it is recommended to use a library where the precursors are annotated with their expected IM. While this is not mandatory, it will positively impact your analysis if your library consists of IM information. To generate a library with IM annotation, you must use the Pulsar search engine as none of the other external search engines like MaxQuant are supported. Pulsar can generate libraries from both PASEF and dia-PASEF runs.

Spectronaut can also generate libraries with in-silico predicted IM. In the library generation settings, you can enable the deep neural network to predict IM in libraries that do not have the IM dimension. This also allows, for instance, to predict IM for libraries generated on other instruments than timsTOF Pro, such as Orbitrap libraries

Library generation	PASEF	dia-PASEF
Pulsar	✓	✓
MaxQuant	✘	✘

2 Getting Started

2.1 Getting Spectronaut®

Spectronaut® software licenses can be requested on our webpage. We also provide free licenses for a trial period upon request on our webpage. After requesting a license, you will get an email with a **link to the installer** and an **activation key** for the software. If this is not the case, please contact us via our Help Center.

IMPORTANT NOTES:

1. Your license will start running from the moment we generate the activation key.
2. Activation keys are computer-bound. If you need to install Spectronaut on more than one computer, please contact us via our Help Center.

Spectronaut Activation

When you install and start Spectronaut for the first time, you will be asked to activate your software by pasting your activation key into the Spectronaut activation dialogue. If your computer can reach our servers, activation will be automatic. If your Spectronaut computer cannot reach our servers (no internet connection, firewall, etc.), you can also activate your software offline. The respective instructions will appear after a few seconds if online activation was not successful: save the registration information file on your computer and send this file to us vie Help Center. In general, you will receive an activation file within one or two working days. To activate Spectronaut using an activation file, click on the "Browse Activation File…" button in the Spectronaut Activation dialogue.

2.2 Demo Data

In Section 3, Spectronaut® Usage, we will guide you through the software perspective by perspective. The examples shown for the classic DIA analysis (Section 3.4.1.1) are generated with the demo data available for downloading here. Please note this demo data was intentionally prepared to be as small as possible for demonstration purposes. Most DIA experiments will require larger storage space and more resources to be analyzed.

3 Spectronaut® Usage

3.1 Structure of Spectronaut®

Layout

Spectronaut® is structured in different levels (Figure 3). The highest level are the perspectives. Within each perspective, you can often find several pages separated into tabs. The layout of each page is normally structured into a left menu (tree) containing elements (nodes) and a right panel containing information related to the selected nodes (plots, reports and summaries). The Analysis perspective features Tree - and a Grid Views.

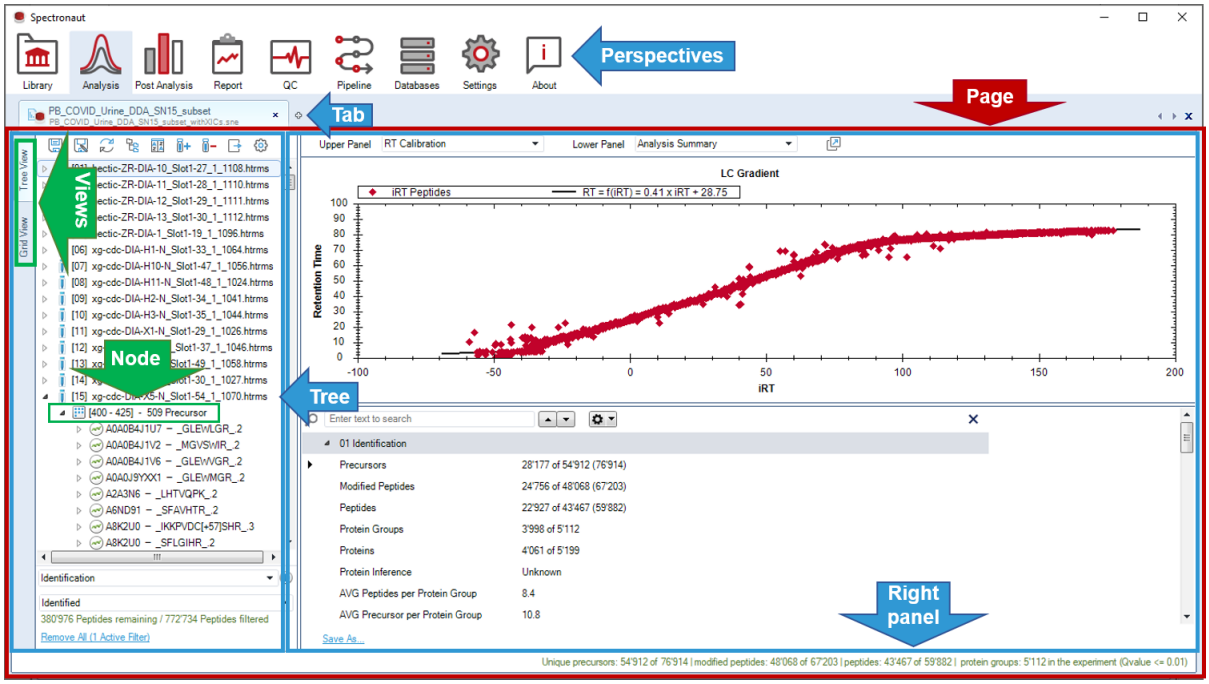


Figure 3. Spectronaut general layout structure.

Tips for a Better Experience

1. Spectronaut is full of informative tool-tips throughout the software (Figure 4). They will appear as you hover over many of the elements.
2. There are many functionalities by right-clicking an element: experiment tab, plots, nodes, etc.

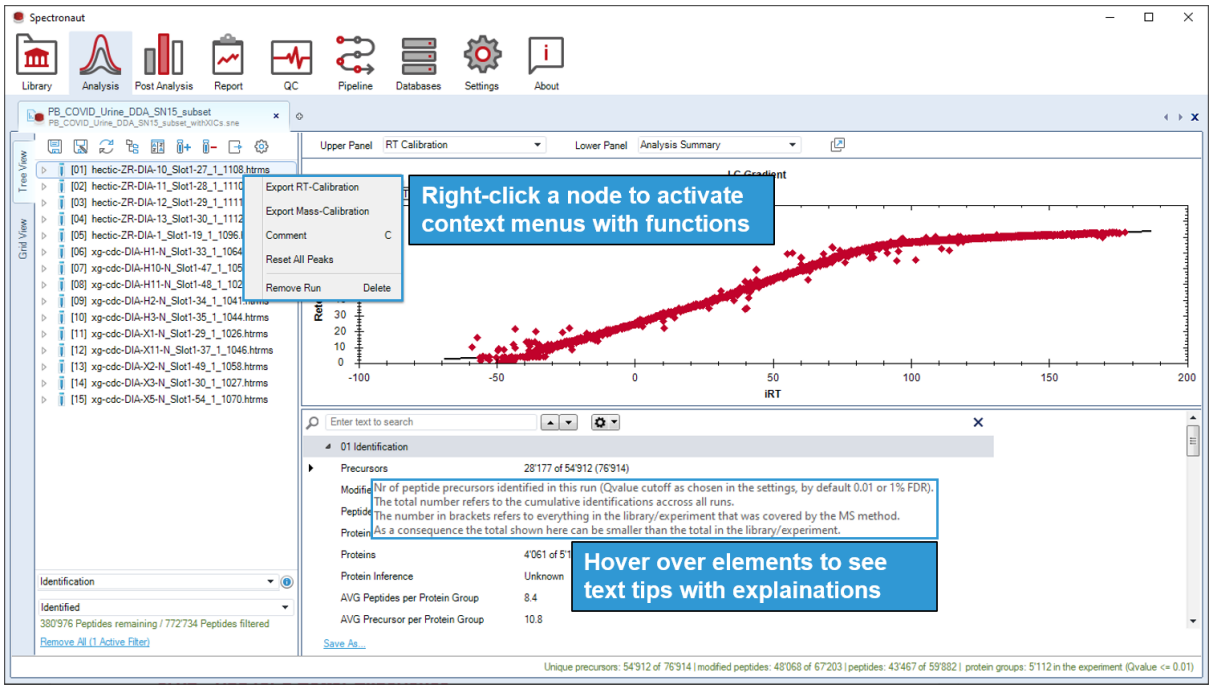
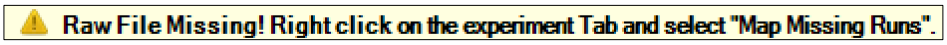


Figure 4. Spectronaut contains nonobvious tips and menus when you hover over some elements or right-click on them.

3. It is possible you get warning messages such as this:



Warnings are sometimes just informational, and do not require action. Errors during spectral library generation and DIA analysis are shown in red. You can display the full library and DIA analysis log in the respective perspective (Figure 5). The log provides messages i.e., information on the steps executed in the pipeline. Warnings and errors are displayed in separate tabs. The log can be directly saved as text file from the Library or Analysis perspective or can be found in the About perspective under "Show error logs". If an error occurs, please send the error log to us via Help Center.

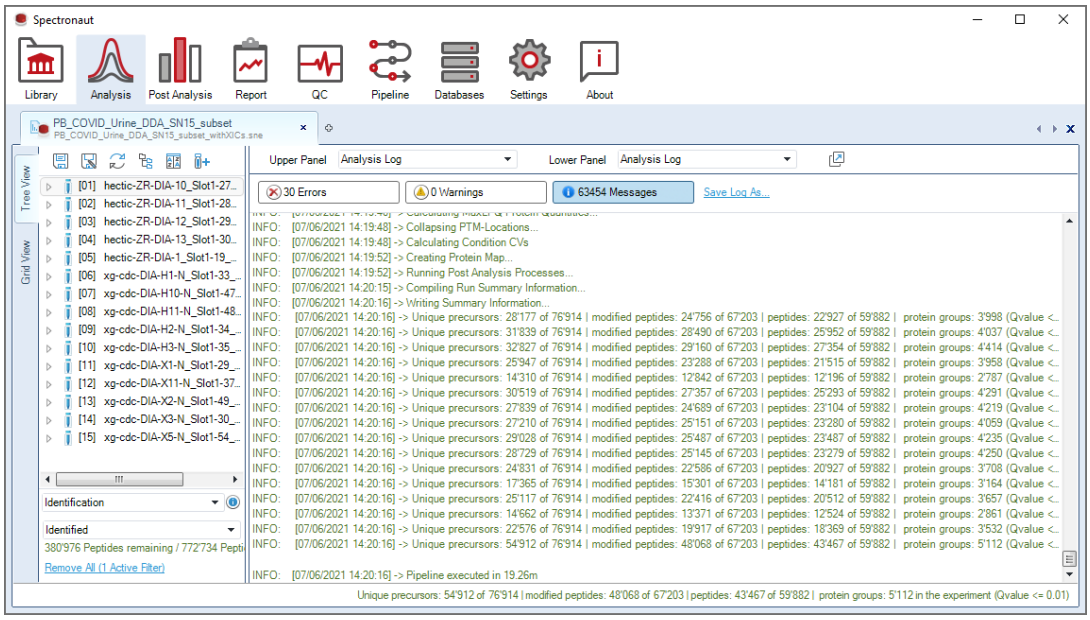


Figure 5. The analysis and library log shows messages, warnings, and errors in separate tabs. One or more tabs can be selected and simultaneouly displayed.

4. Finally, we would also recommend watching Spectronaut video tutorials that you can find here.They will guide you through basic steps of setting up your new DIA analysis with a desired workflow.

3.2 Before Starting

Make sure you have everything you need ready before starting your analysis in Spectronaut®. Two quantitative analyses are supported: classic DIA data extraction using a DDA spectral library or direct search of the DIA data, directDIA™. Table 2 shows which resources are required for each workflow.

Table 2. Input resources for each DIA approach supported in Spectronaut.

Resource	Classic DIA analysis *(library-based)*	directDIA *(library-free)*
DIA run files	required	required [a]
Spectral Library [b]	required	*not applicable*
Protein database (FASTA)	recommended	Required
Gene Ontology annotation	optional	Optional

[a] *directDIA supports Thermo Scientific™, Bruker, and SCIEX run files*

[b] *Spectral Libraries can be generated in the Library Perspective of Spectronaut (Section 0)*

3.3 Library Perspective

The main tasks you can perform in the Library Perspective of Spectronaut® are:

1. Generating a library with Pulsar, Biognosys' proprietary search engine (Section 0).
2. Generating a library using search results from external search engines.
3. Importing an external library.

Guidelines on how to generate the data for an optimal spectral library can be found in Box 4.

Library Generation from Pulsar

Pulsar is Biognosys' proprietary search engine integrated into Spectronaut. We recommend using Pulsar for seamless library generation. Pulsar can search data-dependent acquisition (DDA), data-independent acquisition (DIA), and parallel reaction monitoring (PRM, with MS1 information) data. Both centroid and profile data can be processed. HTRMS-converted runs are also supported. Pulsar is designed to be fast and to scale as the number of runs increases. Further, Pulsar is able to identify co-fragmented peptides in multiple search rounds by subtracting previously identified fragment ions from the spectra. False identifications are controlled by a false discovery rate (FDR) estimation at three levels: peptide-spectrum match (PSM), peptide, and protein group level.

The specific vendors and acquisition modes supported by Pulsar are:

- Thermo Scientific™, with and without FAIMS (DDA, DIA, and PRM with MS1 information)
- SCIEX (DDA and DIA/SWATH™)
- Bruker, including TimsTOF Pro (DDA and DIA)

- Waters (DDA and HDMS^{E})

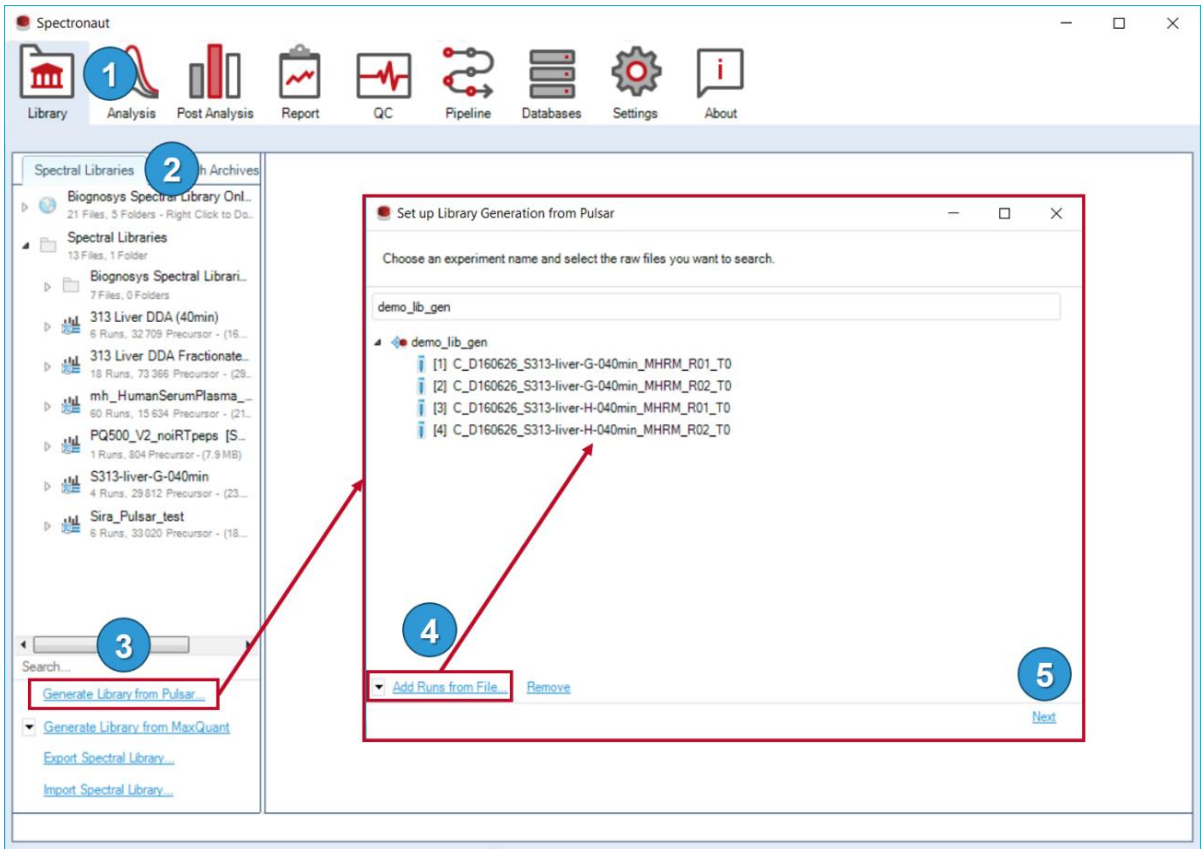


Figure 6. Library Generation with Pulsar. In the Library Perspective, Spectral Library tab click "Generate Library from Pulsar…". Follow the wizard to complete the process.

Every time a search is performed, Spectronaut will save the results (PSMs) for each run as a **Search Archive** (Box 5). These search archives can be used to generate libraries without the need to search the runs from scratch. Run files and Search Archives can be combined conveniently to generate new libraries. Refer to Table 3 to see a summary of the resources you will need in each of these cases.

To generate a library from Pulsar, go to the Library Perspective and click on "Generate Library from Pulsar…" in the bottom left corner. A wizard will appear to help you set up the experiment (Figure 6). A schematic view showing the wizard steps you will encounter, depending on your input resources, is shown in Table 4. The sequential steps are described below:

1. **Set up Library Generation from Pulsar**.
   - Choose an experiment name

- Click on "Add Runs from File…" or "Add Runs from Folder…" and select the runs from which you want to create the library. You can mix runs acquired in different modes***. If you want to generate a library from Search Archives only, skip this step.***

Table 3. Spectronaut can generate libraries using different input resources

Library based on	FASTA file	Gene Ontology information
Run file(s) only	required	optional
Search Archive(s) only	*not applicable*	optional
Run file(s) and Search Archive(s) combined	required	optional

2. **Choose FASTA File(s)** by clicking "Fasta File…". Protein databases can be assigned on a run basis. Multiple protein databases can be selected. FASTA files can be added at this step by clicking "Import…" in the bottom left corner. ***If you are generating a library from Search Archives only, this page will not be shown.***
3. **Choose Pulsar Search Settings** by clicking "Search settings…" (for detailed explanations about each setting, see Appendix 2. Pulsar Search Settings (Section 7.2).
   Choose either the default schema which can be modified on the fly or a previously saved setting schema. Schemas can be assigned at either the experiment or run level. Only one setting schema can be set per run. When nothing is selected for a run, default settings will be applied. ***If you are generating a library from Search Archives only, this page will not be shown.***

Box 4. Library Generation Guidelines

To generate a spectral library, typically DDA runs of your samples of interest are acquired, searched against a sequence database, and the results condensed into a spectral library. To maximize the coverage, we recommend measuring pools of representative samples that have been fractionated (e.g. using high pH reverse phase fractionation). Technical LC-MS/MS replicates are still recommended due to the semi-stochastic nature of DDA proteomics. The optimal number of pools, fractions and replicates depends on the experimental setup and the complexity of the samples. However, overly large spectral libraries where only a small percentage can be recovered from the data might negatively influence the sensitivity of your analysis.

We strongly recommend generating the library in Spectronaut®. Although for most common samples the iRT Kit is not strictly required, we do recommend spiking the iRT Kit into samples aimed at library generation. Spectronaut® will take care of calculating iRTs for all peptides identified even if the iRT Kit was not used.

4. Next, you can add Search Archives (for more information see Box 5) to your library. Search Archives prevent you from having to research run files if you have already searched them in the past.
5. Next, you can choose the FASTA files that will be used for protein inference. If you want to keep the files that were selected for the search space (as set in a previous page for the runs or the ones that were automatically saved in the archives), you can just skip this step.
6. In Spectronaut you can generate your library with Gene Ontology (GO) annotation information. To select a GO annotation (*.goa) file at this point, you need to have the file already imported in the Databases Perspective. Learn how to do so in Section 3.9.1.4.

Box 5. Search Archives

Before Search Archives were introduced, already searched run files had to be searched again from scratch to include them in a library with other runs and maintain control of the FDR. This resulted in a great amount of time and computational resources which had to be reinvested.

With Search Archives, every time a library is generated using Pulsar, the result from this Pulsar search is saved, and will appear in the Search Archive page of the Library Perspective. Search Archives contain the information from a search before applying any FDR filter. This allows several Search Archives to be combined together or with runs files to generate libraries with a proper, library-wide control of the FDR.

7. The next wizard page contains experiment-wide settings for library generation, such as PSM, peptide, and protein FDR thresholds (for a detailed explanation about each setting, see Appendix 4. Library Generation Settings - Section 7.4).

8. The last page shows an overview of the whole experiment set-up. Clicking "Finish" will start the experiment. Using the "View Live Log…" it is possible to follow the progress of the experiment. As soon as the library is generated, it will appear in the library tree. Libraries with FASTA files assigned are marked with a blue protein icon.

Table 4. Schematic view of the wizard steps during library generation depending on the input resources

Library generation step	Run files only	Run files and Search Archives combined	Search Archives only
Choose an experiment name	**required**	**required**	**required**
Add runs	**required**	**required**	*not applicable*
Choose Fasta Files	**required**	**required**	*not applicable*
Choose Search Settings	**required** *(with default)*	**required** *(with default)*	*not applicable*
Choose Search Archives	*not applicable*	**required**	**required**
Specify Protein Inference FASTA files	**optional**	**optional**	**optional**
Specify Gene Ontology annotation	**optional**	**optional**	**optional**
Choose library settings	**required** *(with default)*	**required** *(with default)*	**required** *(with default)*

Library Generation from External Search Engines

We strongly recommend using Pulsar to generate your libraries. However, Spectronaut also supports generating a library from external search engine results. To do so you will need:

1. The search result files or folders → **required**
2. The run files from which the search was done → **required**

Spectronaut supports search results from:

- MaxQuant (Cox et al., 2011)
- Proteome Discoverer™
- ProteinPilot™
- Mascot™

Table 5 summarizes the type of files or folders needed for each search engine, and whether some actions are required for correct integration of the post-translational modification (PTM) annotations.

In addition to the specific result formats above, Spectronaut also supports results in **mzIdentML** (.MZID) format (containing fragment ion information). Finally, any search results can be reformatted into the **Biognosys (BGS) Generic Format**.

Regarding Ion Mobility data, Spectronaut supports PASEF and dia-PASEF™ spectral libraries generated only in Pulsar whereas FAIMS spectral libraries generated by both Pulsar and Proteome Discoverer™. Currently, no other search engines are supported in Spectronaut for PASEF and dia-PASEF™ or FAIMS spectral library generation. Neither BGS generic format can be used for uploading such library.

To generate a library from search results:

1. Go to the Library Perspective → Spectral Library and click on "Generate Spectral Library from…" in the bottom left corner (Figure 7). Choose your search engine.
2. Navigate to the files or folders containing the search results (see Table 5). Spectronaut will try to map the run files automatically (see Box 6). If it fails to do so, you will have to manually link the files by clicking "Assign Shotgun Files…"
3. Choose your library settings in the Library Settings panel or run under default settings. (for a detailed explanation of each setting, see Appendix 4. Library Generation Settings - Section 7.4).
4. Choose a FASTA file in the FASTA File tab. If your FASTA file is not yet in the tree, you can add it at this point by clicking "Import…" in the bottom left corner.

5. Select your Gene Ontology annotation information in the Gene Annotation panel. You should have your file previously loaded into the Databases Perspective (to learn how to do this, go to Section 0). By clicking "Load", Spectronaut will perform the library generation. Your new library will automatically appear in the Library Perspective upon completion.

Box 6. Mapping run files to search results

Spectronaut will try to map the run files automatically by name matching. First, it will look in your Shotgun Raw Repository (Settings Perspective → Global → Directories). If unsuccessful, it will look in the search results location. If the automatic mapping fails, you will see a red cross

If this is the case, you will have to manually map the runs. Click on "Assign Shotgun Files…" to find the missing runs. You can either navigate to a common directory or browse for your runs individually. After the runs have been found, the red cross will change into a green tick mark

Table 5. Supported search engines and information required by Spectronaut when generating a library from search results

Search engine	Search result files	Peptide modifications	
		Default	Custom
MaxQuant	evidence.txt or msms.txt	Included	Imported (*.xml file) from MaxQuant installation folder (\bin\conf\modifications.xml)
Proteome Discoverer	*.msf for PD 1.4 *.pdResult for PD > 2.0	Included with the search results	
Protein Pilot	MS Excel with the suffix "_FDR"	Importation required: "Unified Modification Catalog.xlsx" located in \ProteinPilot\Help folder in Program Files [a]	
Mascot	*.dat	Download the latest Unimod XML database from www.unimod.org/downloads.html [a]	Add manually (see 3.9.1.2)

[a] *These defaults apply only to upgrades from old versions of Spectronaut. If your first version of Spectronaut was either X, 13.0 or 14.0, no action is required concerning default modifications*

3.3.1.1 Spectral Library Generation from BGS Generic Format

Spectronaut supports generating libraries from the minimal BGS Generic Format. This allows end-users to use their favorite search engine with the aid of a basic script which would convert their search result into the BGS Generic Format. This is a tab separated, plain-text format with defined header where each row represents a PSM. Table 6 shows the information required in this file. The user selects the BGS Generic Format file as well as some corresponding LC-MS raw files. Spectronaut tries to automatically map possible modifications to the internal modification database. If unambiguous mapping is not possible then a UI form will prompt the user to make the link.

Box 7. Source-specific iRT calibration

Source-Specific iRT Calibration is a new feature introduced with Spectronaut®.

In the past, the iRT value of a given peptide in a spectral library was summarized by taking a median across all runs where the peptide was identified. However, when building a spectral library from chromatographically heterogenous data, this can lead to a loss in iRT-precision. A good example of this case would be a situation where you would generate a Hybrid Library using DDA files from public repositories together with your quantitative DIA files.

To improve the targeted extraction of such data, we introduced the concept of Source-Specific iRT Calibration. Spectronaut will generate libraries containing as many iRT values as different sources exist in the dataset. When using this library on a quantitative dataset, Spectronaut will use the iRT calibration from the best source available for each assay.

By using source-specific iRT calibration, you will be able to keep the iRT-precision of project specific data while benefiting from the depth of a large resource dataset. This feature though, is not available for the spectral library generation from BGS Generic Format.

Source-Specific iRT Calibration and Search Archives

Spectronaut will create different iRT sources in the libraries in the following cases:

1. Different acquisition methods: DDA and DIA
2. Different calibration types: linear, fragmented (non-linear), *in silico* (no iRT peptides present)
3. Different library inputs: Search Archive(s) or run files. One source will be created per Search Archive. You can create archives with iRT heterogeneity in mind so that you can conveniently reuse them in the future.

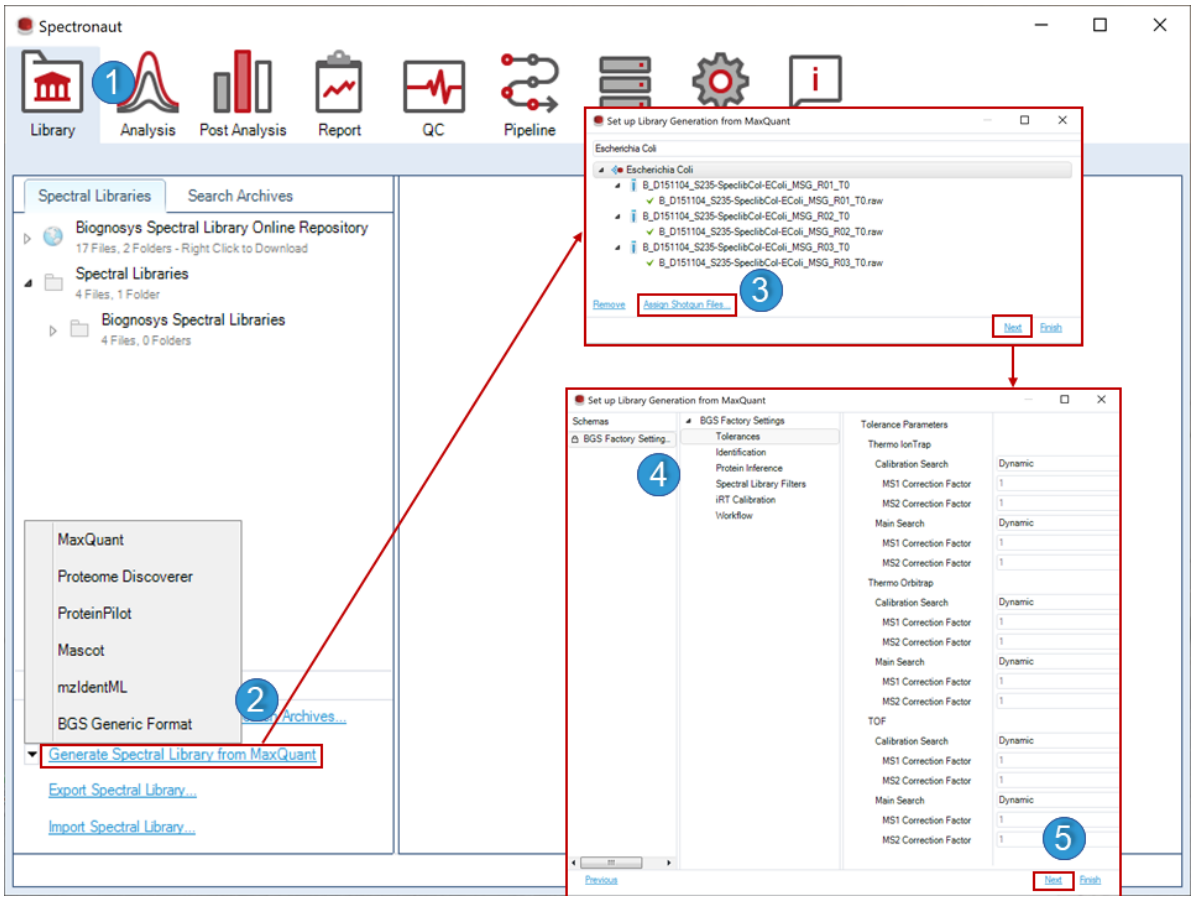


Figure 7. Library Generation from external search engines. In the Library Perspective, Spectral Library tab, click "Generate Library from…" and choose your search engine. Load your search results and assign your run files (Box 6). Follow the wizard to complete the process.

Table 6. BGS Generic Format required information. The BGS Generic Format is a tab-separated, clear text format with a header as specified below. PSM FDR, protein FDR and other filters must be already applied to the PSMs beforehand.

Header	Information	
Raw File	The name of the DDA file in which this PSM was found. This column is used to map the DDA file to the PSM.	**required**
Stripped Sequence	The stripped sequence of the peptide that was found by the search engine for this PSM.	**required**
Precursor Charge	The charge that was associated with this PSM by the search engine.	**required**
Labeled Sequence	The sequence with encoded modification and label information. The only requirement is that the modification/label information should be enclosed in [modification/label] or (modification/label) brackets. Examples: _IHC[CAM]SNYSTQELLR_, _[Ac]AAAAAGGR_, _AAHHPPK_[C-term Mod]	**required**
Retention Time	The retention time of the PSM in minutes if possible. Ideally the apex retention time otherwise the PSM retention time.	**required**
Scan Number	The scan number of the PSM.	**required**
Scan Event	The scan event of the PSM. *This is only relevant for SCIEX *.wiff files.*	**required**
MS1 Intensity	The intensity of the PSM as reported by the search engine.	**recommended**
Protein Group Id	The protein group assigned by the search engine for this peptide. It is not necessary if using protein inference in Spectronaut.	**recommended**

Importing an External Library

To import an external library into Spectronaut, click on "Import Spectral Library…" in the bottom left corner of the Spectral Libraries tab in the Library Perspective (Figure 8).

There are two ways of importing a library into Spectronaut:

1. Importing a *.kit library (Biognosys' library format). In this case, no further action is required, and the library will be loaded automatically into the Library tree.
2. Importing a compatible spreadsheet as a plain text, separated value format (*.txt, *.csv, *.tsv, *.xls). Headers defining your columns are mandatory in these files.

The "Import Spectral Library…" dialogue (Figure 8) will try to auto-detect column names. If there are new column names, Spectronaut will ask you whether or not you want to store them as a recognized synonym for this column. This allows Spectronaut to automatically select these columns the next time you load a spectral library with a similar format (you can remove the user-defined column synonyms in the Databases Perspective → Table Import).

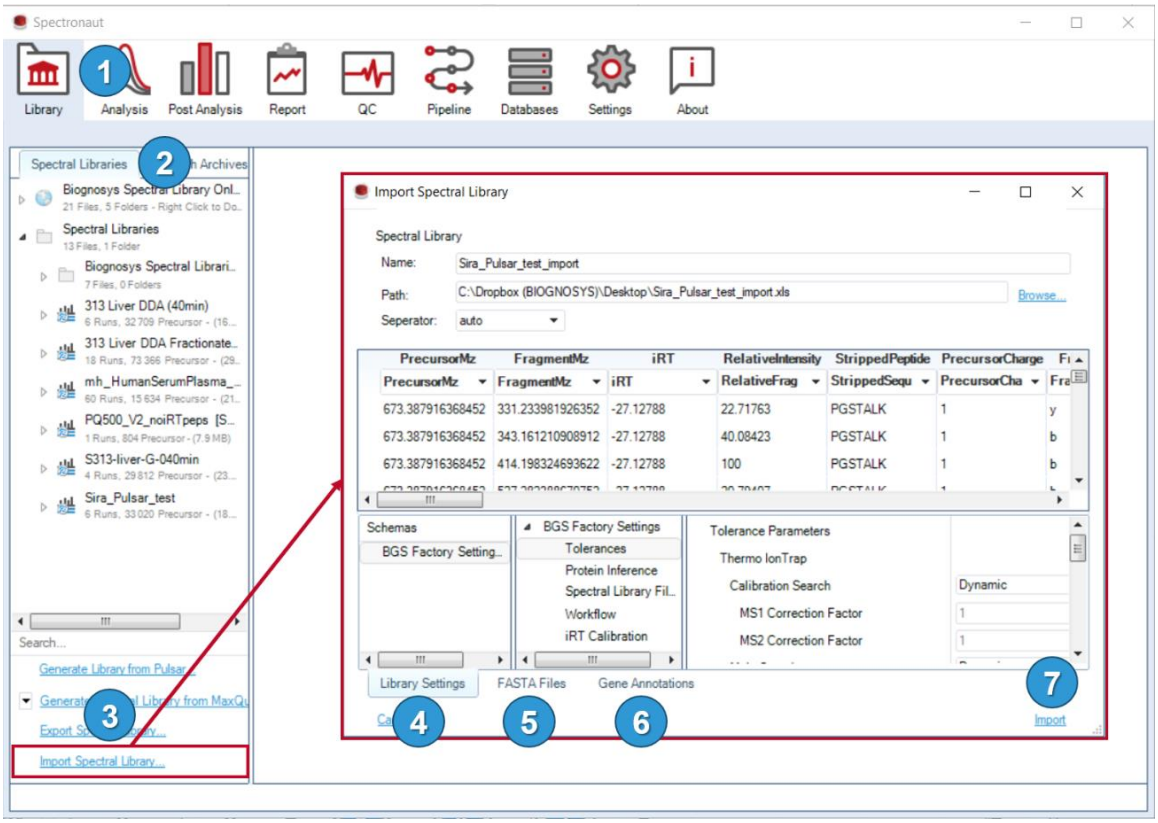


Figure 8. Importing an external library. The "Import Spectral Library…" dialog only applies to formats different than *.kit. You can refine your library using the lower panel tabs.

The import function also allows you to refine your library. In the Library Settings panel, you can choose several options to be applied to your library (for details, see Appendix 4. Library Generation Settings, in Section 7.4). For example, you can perform protein inference again. To do this, go to the FASTA File panel and choose your protein database. You can also add Gene Ontology annotation information using the Gene Annotations panel. Table 7 shows the recommended fields to achieve the best possible results.

3.3.1.2 Library Columns

A spectral library is similar to a typical MRM/SRM transition list. Refer to Table 7 to see what information a library should contain.

Table 7. Library columns to achieve the best possible results with Spectronaut

Header	Requirement	Refers to
PrecursorMz	**Required**	The *in silico* calculated m/z of the peptide precursor ion. Do not round this number.
IonMobility	**optional**	For the FAIMS raw files IonMobility refers to the compensation voltage (CV) applied in the acquisition method. For PASEF data, ion mobility whether empirical or predicted, refers to the value $1/K_0$ expressed in Vs/cm^2, and for $HDMS^E$ files is drift time expressed in ms.
FragmentMz	**Required**	The *in silico* calculated m/z of the peptide fragment ion. Do not round this number.
iRT	**Highly recommended**	The peptide retention time in the reverse phase chromatography converted into iRT space (Escher et al., 2012). If accurate iRTs are provided, the analysis will speed up significantly, and the quality of your results will increase (sensitivity, specificity, accuracy). If no iRTs are available, Spectronaut will predict an iRT for each peptide. This is less accurate than an empirically determined iRT value. To derive iRTs for your peptides, we recommend spiking the Biognosys' iRT Kit into your shotgun runs. If the Library Perspective of Spectronaut is used, iRT values will be automatically determined for your library.
RelativeFragmentIntensity	**Highly recommended**	The relative peptide fragment ion intensity expressed as a percentage of the most intense fragment ion. Please report the raw number without the percentage sign. This information improves the limit of detection.

Header	Requirement	Refers to
StrippedSequence	**Recommended**	The stripped amino acid sequence of the peptide excluding any modifications. Please only use the single letter code for the 20 standard proteinogenic amino acids. This information is used for labeling and scoring your fragment ions in Spectronaut. Further, it is included in the automatically generated unique ID for your precursor if necessary.
PrecursorCharge	**Recommended**	The peptide precursor ion charge. This information is used to label your precursors in Spectronaut and to automatically generate a unique ID for your precursor if necessary.
FragmentType	**Recommended**	The peptide fragment ion type. Usually, this is "y" or "b". This information is used for labeling and scoring your fragment ions in Spectronaut.
FragmentNumber	**Recommended**	The peptide fragment ion number. This number should be between 1 and the length of your peptide in amino acids minus one. This information is used for labeling and scoring your fragment ions in Spectronaut.
FragmentCharge	**Recommended**	The peptide fragment ion charge formatted as a number. This information is used for labeling and scoring your fragment ions in Spectronaut.
FragmentLossType	**Recommended**	Allows you to specify the fragment loss-type (e.g. NH_3 or H_2O). This is mainly used to label your fragment ions within plots.
ExcludeFromQuantification	***If applicable***	This column can be specified optionally and is meant for annotating fragment ions that should only contribute to identification but not to quantification (TRUE or FALSE). TRUE means that Spectronaut will never consider this fragment for quantification and FALSE if Spectronaut will always use it for quantification. If you leave the column empty, the automatic interference correction will decide whether or not to use it (the latter is the default if the column is not specified).

Header	Requirement	Refers to
ModifiedSequence	optional	In the event that your peptide is modified use this column to specify the amino acid sequence including modifications. The modified sequence should be constant for one unique precursor. This information is used to label your precursors in Spectronaut and automatically generate a unique ID if necessary. Spectronaut will try to parse and map modifications from the provided sequences to the internal modification database. This field does not contain any label specific modifications (see LabeledSequence).
LabeledSequence	***If applicable***	Similar to modified sequence. This sequence will specify the respective label for each channel in a labeled experiment. It will also contain any modifications. An example SILAClabeled peptide could look like this: _NAYVC[+57]WTLK_ for the light channel and _NAYVC[+57]WTLK[+8]_ for the heavy channel of the same peptide.
ProteinId	optional	The ID of the protein, the peptide is derived from. This information is used to label your peptides. Spectronaut provides filtering capabilities in the Analysis Perspective including filtering for the protein ID.
EnableForNormalization	optional	Allows to classify precursors to be used for normalization. This can be useful if you know which proteins/peptides constitute a stable background (e.g. in LFQ Benchmark experiments)
UserGroup	optional	A column for additional user information that can be used for grouping and filtering.

To view an example of a library, see our dataset for download or export a library from the Library Perspective in *.tsv format. There is also the option to “Export Peptide List…” by right-clicking on the spectral library of interest. This text file is substantially smaller and more manageable and allows for a quick peek on the peptides contained in the spectral library.

3.3.1.3 Modification Parsing

Once the library is imported, Spectronaut will try to parse all values imported from the "ModifiedSequence" and the "LabeledSequence" columns to assign modification specifications to them. This allows Spectronaut to have greater control over decoy generation. If possible, Spectronaut will automatically assign known modifications from its internal database. If a certain modification is unknown, you will be prompted to assign the modification specification from the database to the new keyword (Figure 9). The only parsing requirement for external modification definitions is the modification tag which is specified within round or square brackets. Spectronaut will not parse modifications specified as single letter special amino acids (such as 'B' for carbamidomethyl cysteine or 'O' for oxidized methionine). You can remove previously assigned parsing synonyms in the Modifications page of the Databases Perspective (see Section 0).

Merging Libraries

Although Spectronaut allows merging libraries, we do not recommend this action, since this can lead to uncontrolled inflation of the protein FDR. Instead, we strongly advise generating a new library from a search that includes all relevant LC-MS/MS runs, so FDR remains controlled. Alternatively, you can use Pulsar Search Archives (see Section 0 and Box 5). This being said, two or more spectral libraries can be merged in the Library Perspective. To do so, select the libraries you would like to merge while holding the **Ctrl** key and then, right-click to open the context menu and select the "Merge" option. This will open a setup window similar to when generating a library from a database search (see Figure 7).

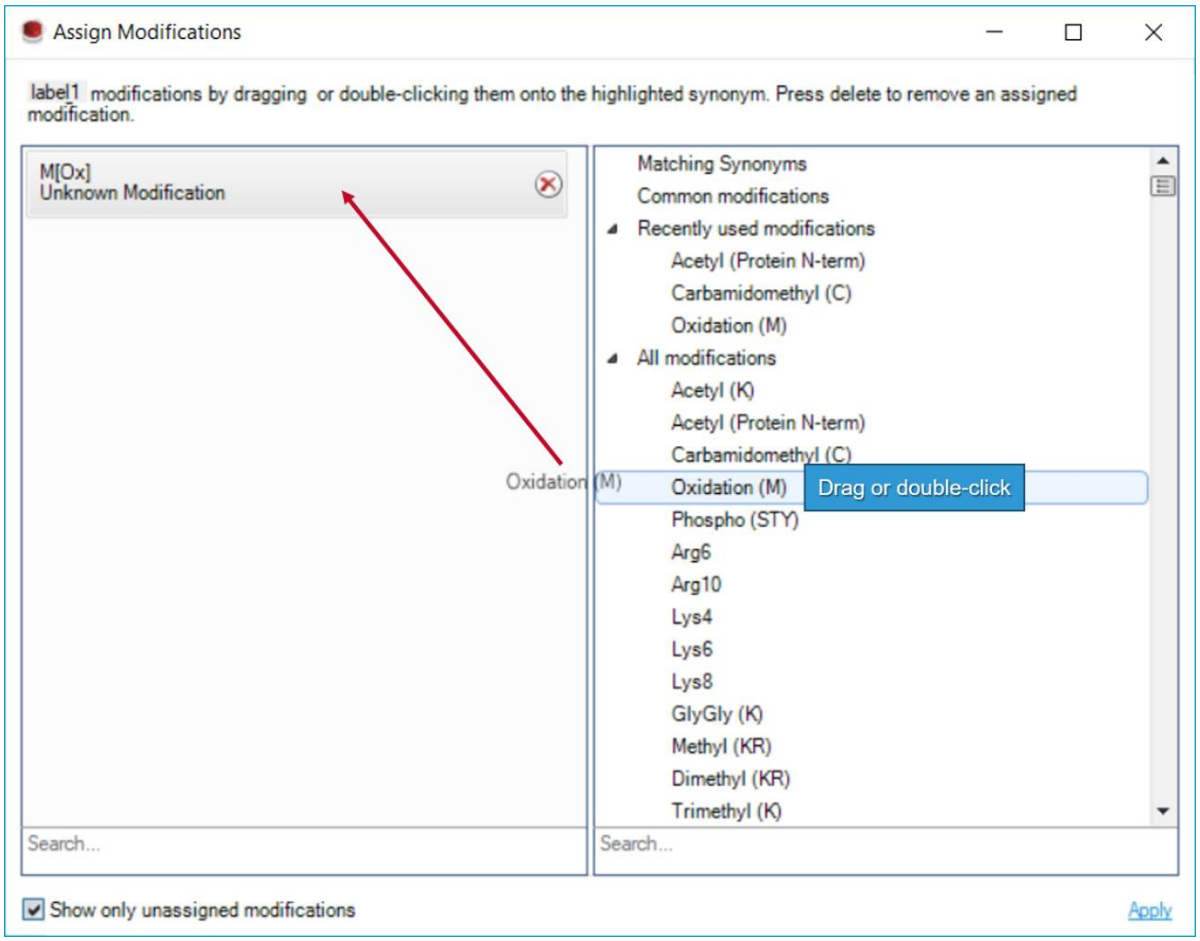


Figure 9. Modification assignment during import of an external library. Add the synonym from the database by double-clicking of dragging it to the unassigned modifications.

Please note that if the libraries used for merging have different types of protein annotations, protein counts in the merged library will be inflated (as the same protein could be counted twice). This will not happen if the libraries were generated in Spectronaut performing protein inference using the integrated IDPicker algorithm (Zhang et al., 2007).

Library Overview

Spectronaut provides several different plots with an overview of your library. You can access these plots by clicking on the library node in the tree and selecting the relevant plot in the right panel (Figure 10).

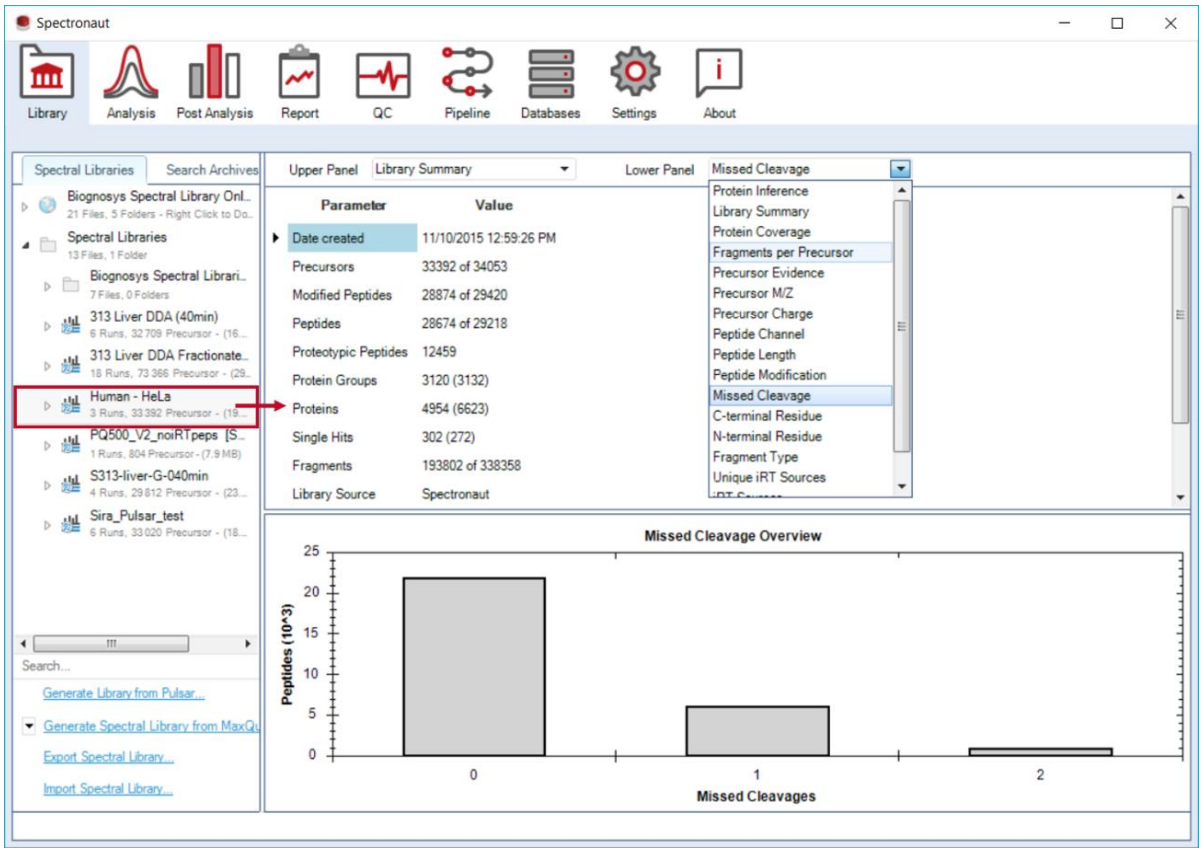


Figure 10. Spectral library overview. Several plots can be selected from the drop-down menus. In this example, the top plot shows the Library Summary, while the bottom plot shows the missed cleavage overview

Making a Labeled or a Spike-in Library

There are two ways to generate a labeled (or spike-in) library. Either from an existing label-free library or from scratch. If you already have a label-free library that you would like to label, you can in silico generate a labeled library from a label-free library. By right-clicking on a library in the Library Perspective, you can attach heavy labels to an existing library (Generate Labeled Library). Doing so will open the Label Editor form where you can select which labels should be applied to the existing library by double-clicking from the list on the right-hand side (Figure 11). The selected library will be stripped of any pre-existing labels. The selected workflow will be included in the library to define how these new peptides will be treated during analysis.

If you want to generate a labeled library from scratch, specify the channels (up to three channels allowed) and labels to be searched (Pulsar Search > Labeling > Labeling applied, and tick the box). Note that also the light channel has to be specified by selecting a channel and leaving the labels text box empty. Then under the Workflow node of the Library Generation settings, select “In-Silico Generate Missing Channels” and choose the appropriate workflow

(i.e., label, spike-in, inverted spike-in). This option will add the missing channels to yield a homogeneous set of channels for all peptides.

For detailed information about the supported, labeled workflows, see Section 3.4.1.6. Additionally, find a short video tutorial on labeled library generation here.

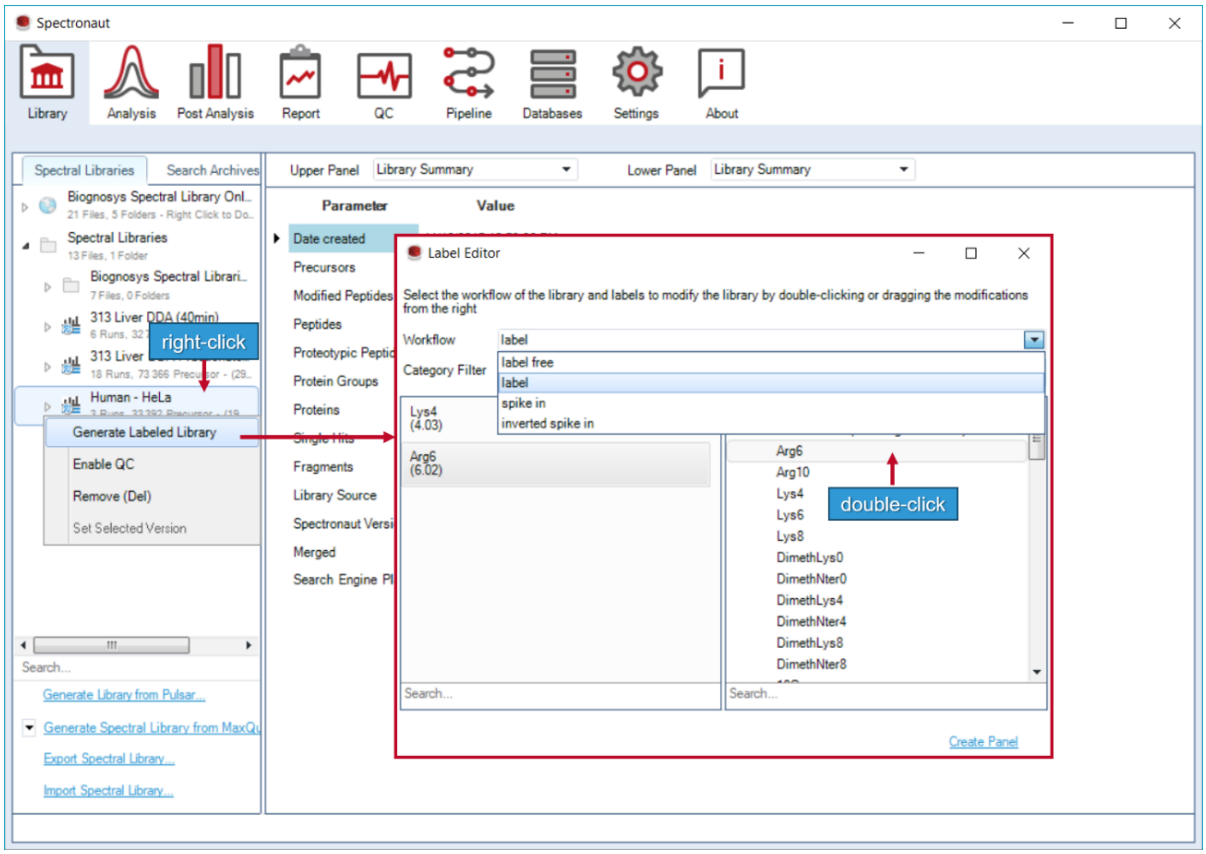


Figure 11. Right-clicking on a Library opens a context menu with several options, such as generating a labeled library or enable library for QC. In the figure there is an example of applying a SILAC label to an existing library. The two isotopic modifications Arg6 and Lys4 are selected by double-clicking to be applied as label to all applicable peptides.

Generate a Quality Control Kit from a Library

When right-clicking on a spectral library in the Library Perspective you have the option to generate a new QC kit using this library (Enable QC, Figure 11). This will select 250 highly abundant peptides from the spectral library which will be added as a QC kit to the quality control perspective. The selection of peptides can also be altered manually within the dialogue. These peptides can then be tracked for quality control purposes within the quality control perspective whenever the corresponding library was used.

3.4 Analysis Perspective

Spectronaut® starts up in the Analysis Perspective. This perspective allows you to:

1. Set up a classic DIA or a directDIA™ analysis
2. Load an existing experiment (SNE file)
3. Review your data at any level through many of the useful plots, reports, and filters
4. Refine your quantification by modifying integration boundaries, fragment ions used for quantification, defining interferences, etc.

Setting up and Running an Analysis

Setting up your DIA analysis is straightforward, thanks to the setup wizards in Spectronaut.

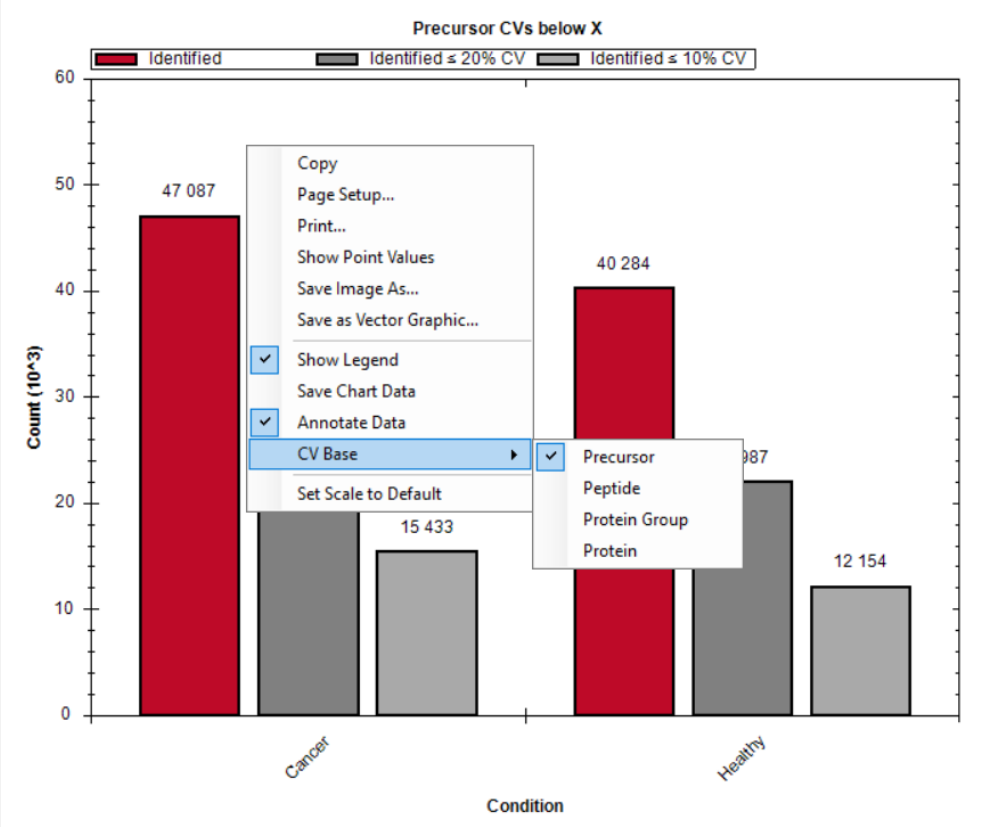
Before starting, see Table 2 to make sure you have everything you need. After completing the wizard and clicking "Finish", Spectronaut will switch back to the Analysis Perspective and start the analysis.

First, in every run, Spectronaut will perform a basic, linear iRT calibration using the iRT Kit peptides; then, Precision iRT calibration will be applied using a stored set of endogenous iRT peptides or using the novel Biognosys' Deep Learning Assisted iRT Calibration (as defined in the settings). Browsing your data is possible a few seconds after the initial calibration process is finished. Once the analysis has finished the number of unique precursors, peptides and protein groups identified for the q-value cutoff defined in the settings (default 0.01, the equivalent of an FDR cutoff of 1%) will be shown at the bottom right (Figure 22).

Box 8. Spectronaut plots: how to get the most out of them

Spectronaut provides, across all its perspectives, a comprehensive number of plots of many types to show you all the relevant details about your analysis, from MS data acquisition to post-analysis results. Most plots in Spectronaut are interactive and customizable to some extend. For example: zoom in on a plot by selecting the area you want to enlarge (find back to default scale by right-clicking on the plot). Drag or navigate a plot horizontally by Ctrl +click and drag.

By right-clicking on a plot, you will find a context menu with an extensive list of functionalities (see figure below):



In this example, you can show or hide the legend, save the data used for the plot, choose the unit you want to show numbers for (protein, peptides), among many others.

3.4.1.1 Performing a DIA Library-based Analysis

Most examples shown in this Section are generated with the available dataset for download. Please, use this data to reproduce these results. Alternatively, you can generate your own DIA data to test Spectronaut. To check what resources you will need to perform a DIA library-based analysis, refer to Table 2.

To start a library-based DIA analysis, go to the Analysis Perspective (add a new experiment tab if needed) and click on "Set up a DIA Analysis from…" in the bottom left corner. This will

let you navigate to your run files or folders. Another option is to copy the run files from a folder to the clipboard. Spectronaut will auto-transfer the runs saved in the clipboard when setting up an analysis. Once you have chosen your DIA data, a wizard will start guiding you through the set-up (Figure 12):

1. **Set up a DIA Analysis**.
   - Give a name to your analysis.
   - Select the whole experiment or individual runs, then click on "Assign Spectral Library…". Different libraries can be assigned to different runs. More than one library can be assigned.

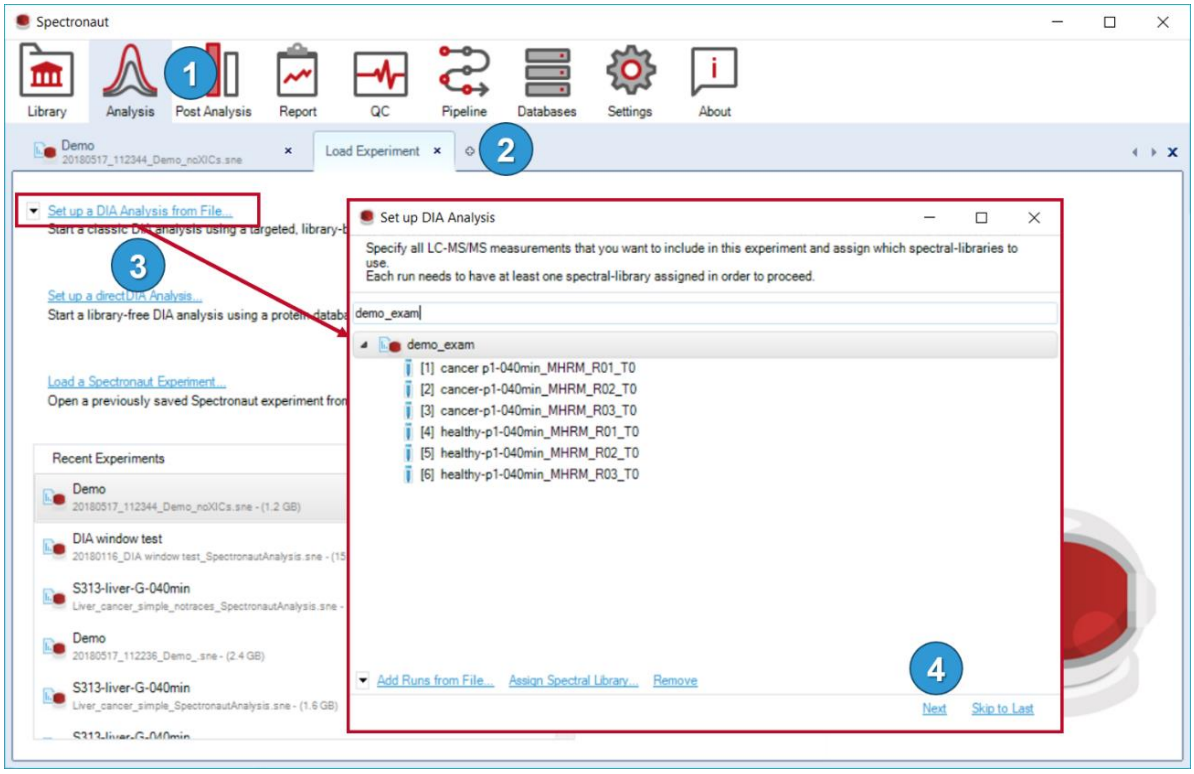


Figure 12. Setting up a library-based DIA analysis. After selecting your run files, a wizard will guide you through the process. You will be prompted to assign one or more libraries, the FASTA files, the GO annotation file, and to select your analysis settings. Finally, you will see a summary of the analysis set-up (Figure 13).

2. **Choose Spectral Library.** Select the library from the Recently Used list, From File, or From Library Perspective. If the library is chosen from file, this action is similar to importing an external library, described in Section 0. Click "Load" to add the library to your analysis.
3. On the next page, you select your DIA Analysis Settings Schema. Use one of the schemas available or modify one on the fly. These settings will define many important aspects of the analysis, such as FDR thresholds, quantification preferences, how to filter your data,

among others. The BGS Factory Settings (default) schema is a good starting point for most projects. Find a detailed explanation of the analysis settings in Appendix 1. DIA Analysis Settings (Section 7.1).

4. Choose a protein database (FASTA file) if you want Spectronaut to perform protein inference. Spectronaut performs protein inference according to the IDPicker algorithm (Zhang et al., 2007). Refer to Appendix 1. DIA Analysis Settings (Section 7.1) for more details about this option.
5. Specify your experimental set-up (conditions, replicates) so Spectronaut can test for differential abundance (paired and unpaired Student's t-test) and perform other Post Analysis processing steps. See Section 3.4.1.5 for more information about how the condition editor works.
6. Choose a Gene Ontology (GO) annotation file if you want Spectronaut to give you extra biological insight into your experiment. This includes GO term enrichment and GO clustering.
7. Before clicking "Finish", a summary of your analysis set-up will be shown (Figure 13).

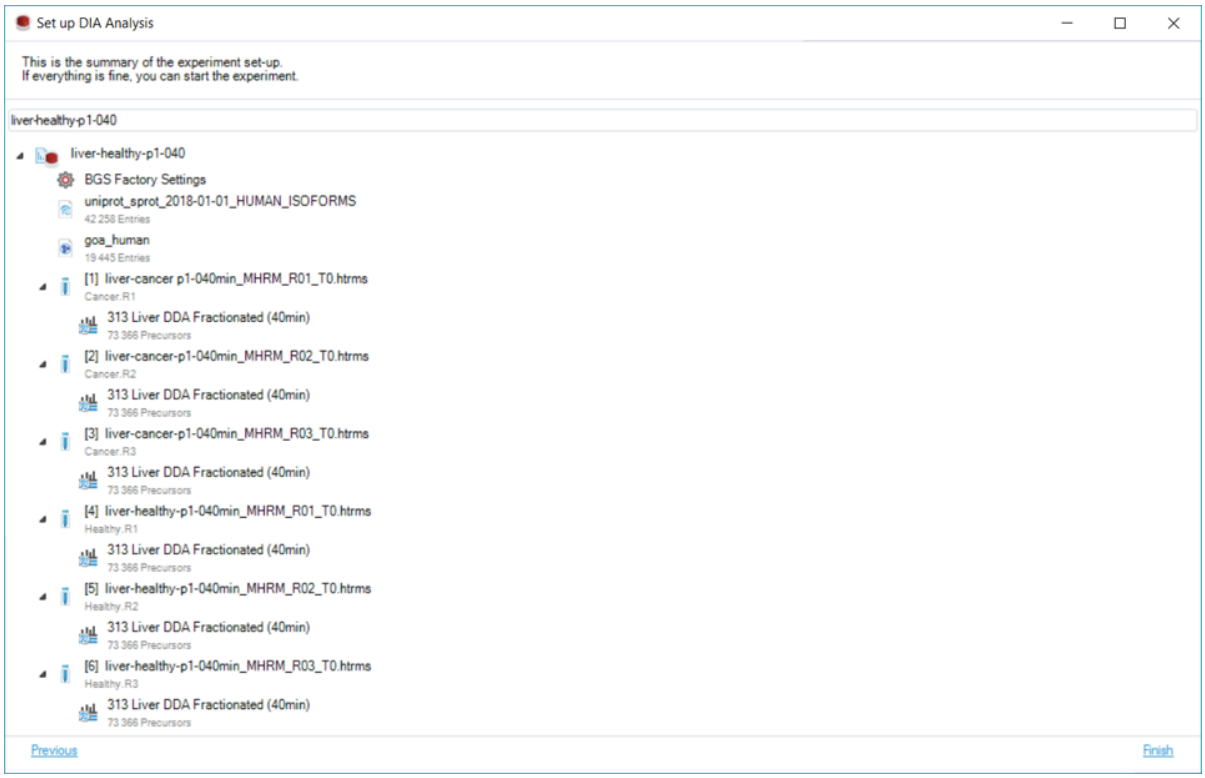


Figure 13. Summary of the analysis set-up. Click "Finish" to proceed with the calculations.

3.4.1.2 Performing a directDIA™ Analysis

Spectronaut enables directDIA™, Biognosys' library-free DIA workflow. This novel workflow allows you to directly search DIA files using nothing but a FASTA file for identification. DirectDIA is a 2-step process: 1) It creates a library by performing a spectrum-centric analysis of your DIA runs based on the specified protein database which is conceptually similar to DIA Umpire (Tsou et al., 2015), and 2) It automatically uses the library to perform a targeted analysis of the same DIA runs. In the end, you get the results from the targeted analysis and optionally, you can save the generated search archive by selecting the option in the Reporting node of the Global Settings. Currently, directDIA is supported for Thermo Scientific™ Orbitrap, SCIEX TripleTOF® data, Bruker timsTOF Pro, and Waters Synapt data. To know the resources required to perform a directDIA analysis, refer to Table 2.

To start a new directDIA analysis, go to the Analysis Perspective (add a new experiment tab if needed) and click "Set up a directDIA Analysis…" in the bottom-left corner. This will prompt you to navigate to your run files. Another option is to copy the run files from a folder to the clipboard. Spectronaut will auto-transfer the runs saved in the clipboard when setting up an analysis. Once you have chosen your DIA data, a wizard will start guiding you through the set-up (Figure 14):

1. Set up a directDIA Analysis. Give your analysis a proper name.
2. Choose one or more FASTA files. The searches will be done against these databases.
3. Choose the search and analysis settings for your experiment. Use one of the schemas available or modify one on the fly. These settings will define many important aspects of the analysis, such as digestion rules, FDR thresholds, quantification preferences, how to filter your data, etc. The BGS Factory Settings (default) schema is fitting most projects. Find a detailed explanation of the directDIA settings see Appendix 3. directDIA™ Settings (Section 7.3).
4. Specify your experimental set-up (conditions, replicates) so Spectronaut can test for differential abundance and perform other Post Analysis processing steps. See Section 3.4.1.5 for more information on how the condition editor works.
5. Choose the Gene Ontology (GO) annotation if you want Spectronaut to provide extra biological insight for your experiment, including GO term enrichment and GO clustering.
6. Optionally, additional run files (DDA and DIA) or Search Archives can be added to create a hybrid library. Those runs/Search Archives will be used to enrich the project-specific library and improve the depth of proteome coverage in the analyzed DIA project.

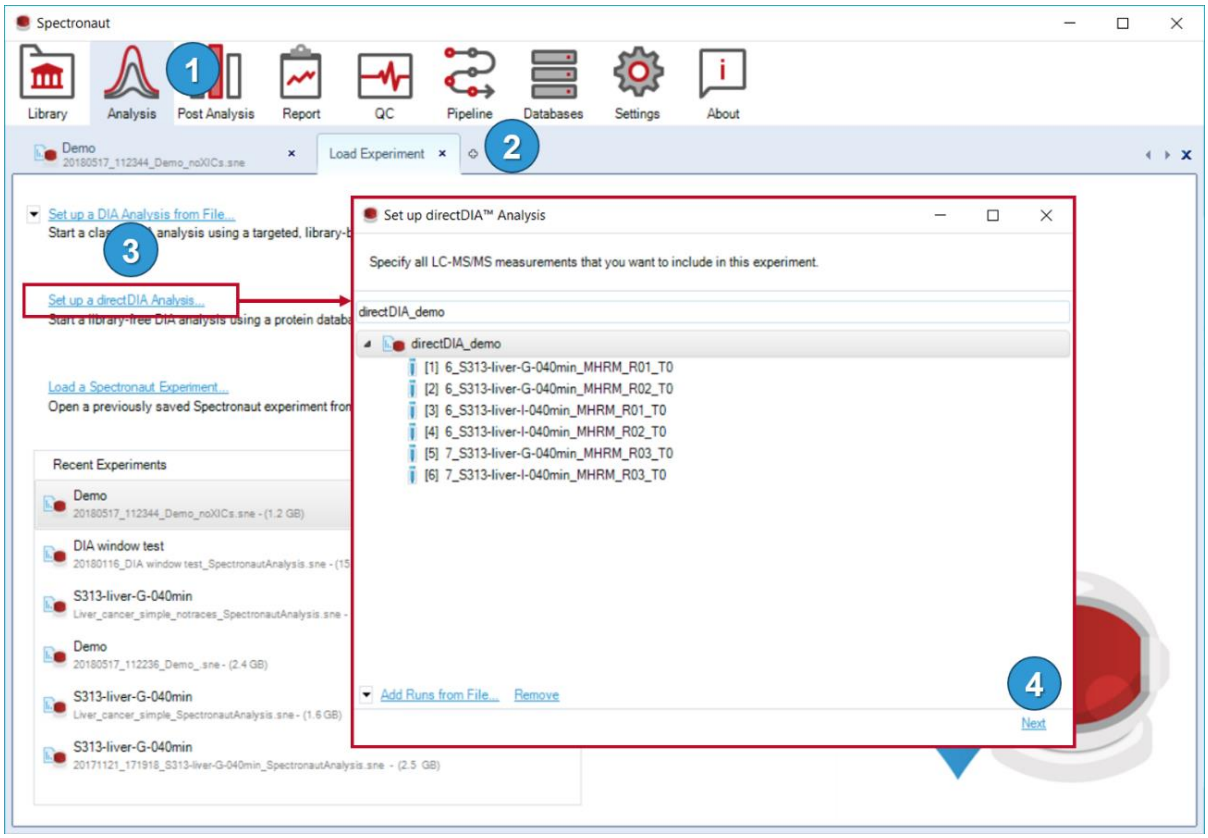


Figure 14. Starting a directDIA analysis. After selecting your run files, a wizard will guide you through the process. You will be prompted to assign the FASTA files, the GO annotation file, and to select your analysis settings. Finally, you will see a summary of the analysis set-up before clicking "Finish".

3.4.1.3 Method Evaluation Workflow for directDIA

This feature allows for the direct comparison of different DIA methods within one directDIA Experiment. By enabling the Method Evaluation parameter in the directDIA Worklow settings (Figure 15), the user will perform a separate Pulsar DIA search per condition to better compare different DIA methods within one experiment. The Method Evaluation worklow is not suited for quantittave experiments.

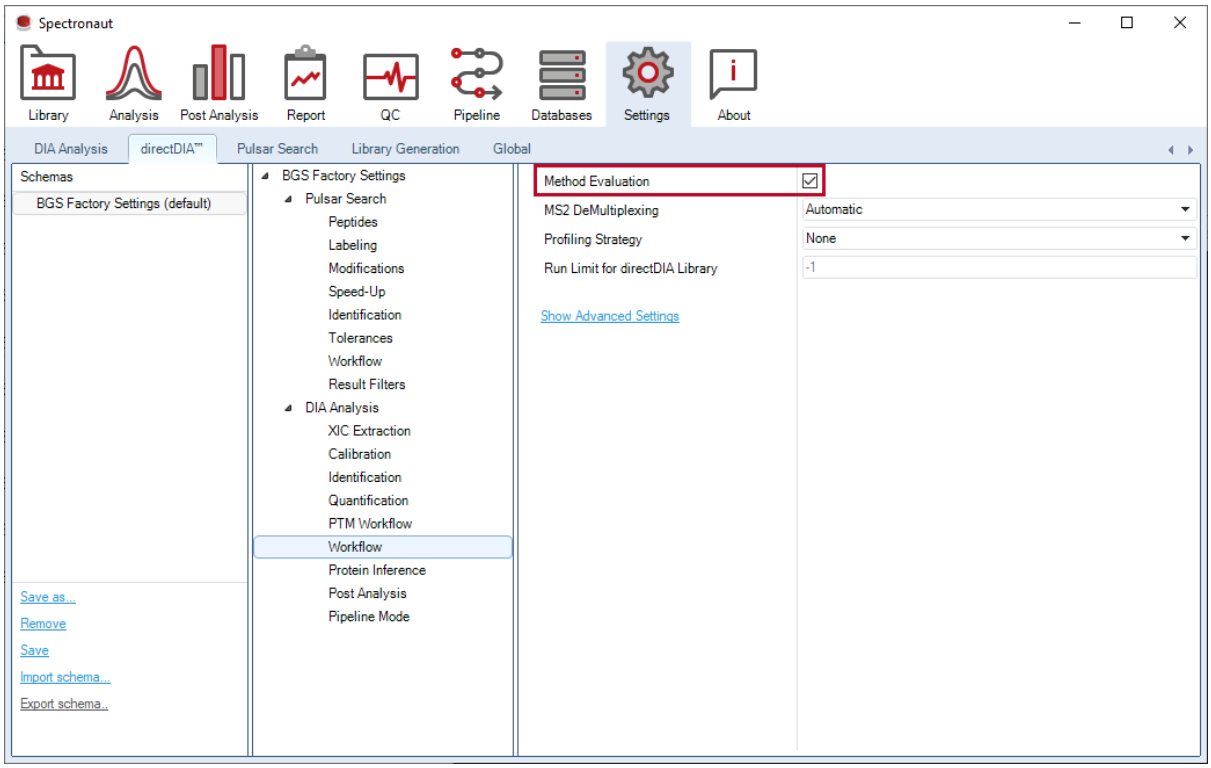


Figure 15. The Method Evaluation workflow allows for direct comparison of different DIA methods within one directDIA experiment.

3.4.1.4 directDIA+

Spectronaut 17 introduced the new directDIA+ workflow for significantly deeper coverage from our library free directDIA pipeline. By default, directDIA+ is always enabled but there is the option to select from three different workflows (Figure 16).

- **directDIA**

  This will select the classical directDIA pipeline without the directDIA+ extension. This workflow is still well suited for many-modification search spaces while being significantly faster then directDIA+ (Deep). It will however, yield significantly less identifications in most cases.

- **directDIA+ (Deep)**

  This option will always provide the deepest coverage in all cases. However, the relative gain between directDIA and directDIA+ (Deep) for search spaces with many modifications can be smaller.

- **directDIA+ (Fast)**

  This will give you a good balance between speed and depths for classical (non-PTM) search spaces. While this option will be the fastest one, it might yield less

identifications for large PTM search spaces compared to directDIA+ (Deep) or the classical directDIA workflow.

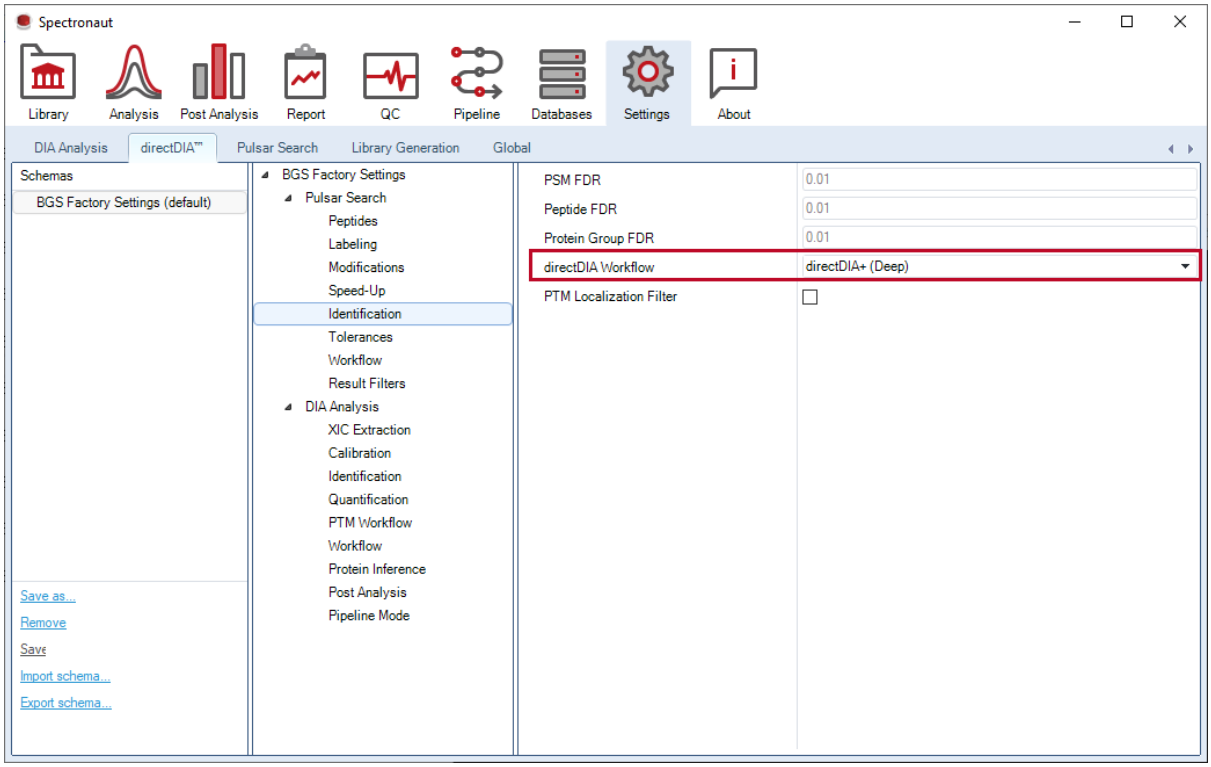


Figure 16. The workflow option for directDIA+ allows you to prioritize speed over depth.

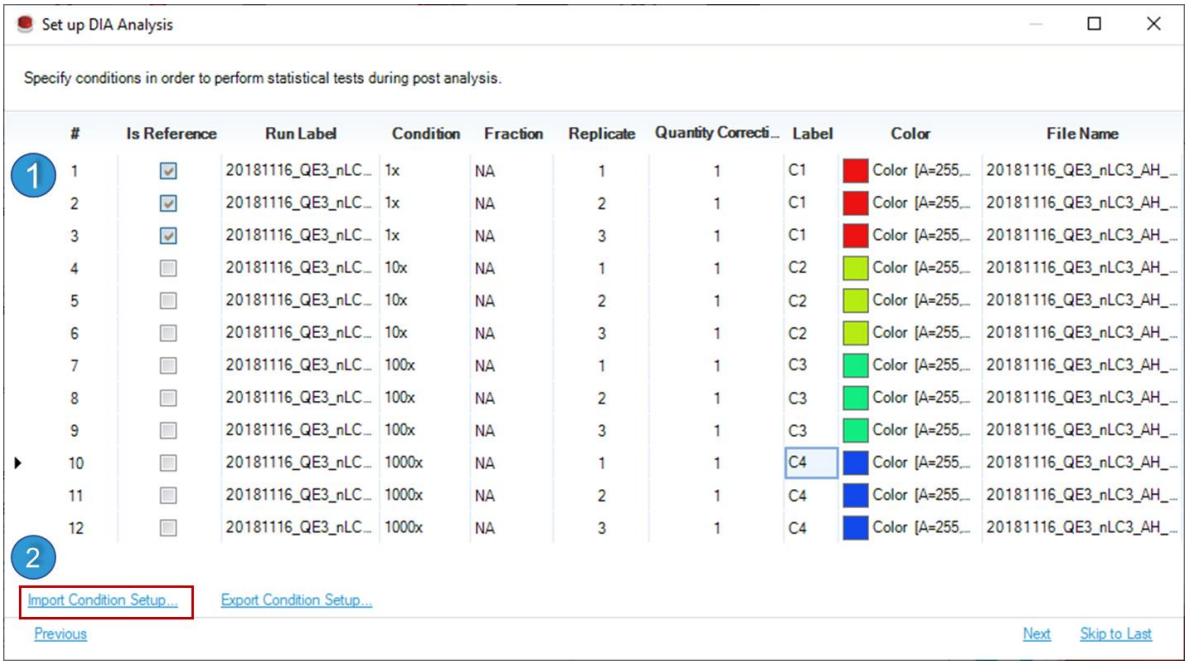
3.4.1.5 Configure Conditions

To let Spectronaut perform the differential abundance tests (paired or unpaired Student's t-test) and other condition-wise metrics, you need to specify your experimental set-up during the configuration of your analysis. Spectronaut will ask you to annotate your runs and specify to which condition, biological replicate, and fraction (if applicable) they belong (see Box 9 to learn more about fractionation in Spectronaut). Each condition in Spectronaut will get a color assigned during the set-up which will be used for post-analysis plot labelling. The Condition Set up panel contains several columns (Figure 17):

- The "Label" column is used for plotting purposes.
- You can specify which condition should be considered as the reference ("Is Reference") for differential abundance testing, and all other conditions will be compared against that one.

You can always redefine and rerun the comparisons in the Analysis perspective under Settings, Condition Setup.

- You also have the option to enter quantity correction factors for each sample. Spectronaut will multiply the final quantities by this factor. An application example would be if quantities should be expressed per initial sample volume (for instance in plasma).



#	Is Reference	Run Label	Condition	Fraction	Replicate	Quantity Correcti...	Label	Color	File Name
1	☑	20181116_QE3_nLC...	1x	NA	1	1	C1	Color [A=255,...	20181116_QE3_nLC3_AH_...
2	☑	20181116_QE3_nLC...	1x	NA	2	1	C1	Color [A=255,...	20181116_QE3_nLC3_AH_...
3	☑	20181116_QE3_nLC...	1x	NA	3	1	C1	Color [A=255,...	20181116_QE3_nLC3_AH_...
4	☐	20181116_QE3_nLC...	10x	NA	1	1	C2	Color [A=255,...	20181116_QE3_nLC3_AH_...
5	☐	20181116_QE3_nLC...	10x	NA	2	1	C2	Color [A=255,...	20181116_QE3_nLC3_AH_...
6	☐	20181116_QE3_nLC...	10x	NA	3	1	C2	Color [A=255,...	20181116_QE3_nLC3_AH_...
7	☐	20181116_QE3_nLC...	100x	NA	1	1	C3	Color [A=255,...	20181116_QE3_nLC3_AH_...
8	☐	20181116_QE3_nLC...	100x	NA	2	1	C3	Color [A=255,...	20181116_QE3_nLC3_AH_...
9	☐	20181116_QE3_nLC...	100x	NA	3	1	C3	Color [A=255,...	20181116_QE3_nLC3_AH_...
10	☐	20181116_QE3_nLC...	1000x	NA	1	1	C4	Color [A=255,...	20181116_QE3_nLC3_AH_...
11	☐	20181116_QE3_nLC...	1000x	NA	2	1	C4	Color [A=255,...	20181116_QE3_nLC3_AH_...
12	☐	20181116_QE3_nLC...	1000x	NA	3	1	C4	Color [A=255,...	20181116_QE3_nLC3_AH_...

Figure 17. Condition Setup panel during DIA Analysis set-up. You can manually adjust your conditions on the panel or Import Condition Setup from a text file.

Unless actively changed to pairwise comparison (paired Student's t-test), or disabled in the Analysis Settings, Spectronaut will perform an unpaired comparison (two-sample t-test) between all conditions specified in the Condition Setup panel. The results are reported in the Post Analysis Perspective.

There are several ways to introduce the annotation information into the Condition Setup panel:

1. If you maintain a file-name structure which is self-annotating, you can define a parsing rule to automatically parse the conditions and replicates from it (Settings → Global → General → File Name Parsing Schema; see Section 3.10.1.1). A parsing rule is a set of instructions that inform Spectronaut what type of information you want to extract from the file name and how.

2. The Condition Setup table is editable: you can directly write in any of the fields to enter your information (Figure 17). The table will recognize your changes and adapt them to the rest of the fields automatically. Be aware that the condition editor is space and case sensitive.
3. Import your annotation from an external text file. The easiest way to do this is by exporting the Conditions set-up Spectronaut suggests, modify it, save it as a text file and import it back.

Box 9. Sample Fractionation in Spectronaut

We do not recommend sample fractionation in DIA analyses. While for DDA, sample fractionation results in significantly higher coverage, the gains in the case of DIA are less significant. In general, increasing the coverage in a DIA analysis is achieved by optimizing the acquisition method and building a better spectral library.

In addition, one of the main features of DIA datasets is the low CVs and the high reproducibility. The process of sample fractionation introduces variability, sometimes notably high, which renders in a dataset of lower quality.

Although not recommended, Spectronaut supports sample fractionation. If you have your samples fractionated, you need to annotate this properly in the Condition Setup. This will allow Spectronaut to perform fraction-wise normalization. Furthermore, libraries might have to be optimized in case of fractionation. If a peptide is not expected to be in a fraction, ideally it should not be targeted in that fraction. The reason for this is shown in the following table:

Peptide present only in Fraction1	Identified in Fraction1 Replica1	Identified in Fraction1 Replica2	Identified in Fraction2 Replica1	Identified in Fraction2 Replica2	**Full Profile**	**Sparse Profile**
A	TRUE	TRUE	FALSE	FALSE	**FALSE**	**TRUE**
B	TRUE	FALSE	Not Targeted	Not Targeted	**FALSE**	**TRUE**
C	TRUE	TRUE	Not Targeted	Not Targeted	**TRUE**	**TRUE**

3.4.1.6 Workflows Supported in Spectronaut

Labeled Workflows

Spectronaut not only performs label-free quantification. Labeled workflows are also supported, and specific scoring methods are developed for each approach.

- **Label-free**: Default workflow for all channel experiments. Peak detection, scoring and identification are applied as usual.
- **Labeled**: Peak detection and scoring will be applied to all channels. Quantification in Post Analysis will be performed on the light to heavy ratio.
- **Spike-in**: Peak detection will be performed on only the reference (heavy) channel. Scoring and identification will be performed on the target (light) channel. The heavy channel is expected to be easily detectable and considered a peak-picking aid in this experiment. Quantification in Post Analysis will be performed on the target to reference ratio.
- **Inverted Spike-in**: Similar to spike-in, but the light channel is considered as the reference.

For more details about how to set these workflows, see Appendix 1. DIA Analysis Settings (Section 7.1).

Please note that for labeled workflows post analysis with statistical testing is not currently supported in Spectronaut. Both those steps could be done in downstream process after exporting Spectronaut Analysis Report. All the labeled workflows reports contain separate columns showing light, medium, and heavy channels as well as ratio of target to reference.

Spectronaut also supports the analysis of raw data acquired with Ion Mobility dimension. You can find more information on supported ion mobility DIA acquisition methods in Box 2 and Box 3 (Section 1.5). Ion Mobility acquisition is compatible and could be used for implementation of all supported workflows.

PTM Workflow

Spectronaut features PTM workflow dedicated to determination of posttranslational modifications occurrence, confidence of their localization and their differential abundance analysis (Bekker-Jensen, Bernhardt, et al., 2020). For the details how the PTM localization is performed, see Box 10, an example of PTM localization plot could be found on Figure 63.

The differential abundance testing is available at the modification site level if the PTM analysis is selected.

Firstly, Spectronaut performs quantitative site collapse of parent peptides carrying given modification at a specific modification site (Bekker-Jensen, 2020). If the parent peptides carry more modifications of the same type, a separate collapse could be performed according to the modification multiplicity (see collapse of doubly and singly phosphorylated peptides on Figure 18). Singly phosphorylated parent peptides (multiplicity 1, M1) and doubly phosphorylated (multiplicity 2, M2) will undergo site collapse separately. If the peptide carries three or more modifications of the same type, it will be reported with multiplicity 3, M3.

Subsequently, resulting modification site objects undergo differential abundance analysis.

Multiplicity On	Modification site Quant
PX_T10_M1	300
PX_T10_M2	200
PX_S15_M2	200

Figure 18. Example of quantitative site collapse of phosphorylated parent peptides, performed according to their multiplicity.

Box 10. PTM localization algorithm

A major challenge for library based (peptide-centric) data independent acquisition (DIA) is the correct localization of post translational modifications (PTMs). Standard targeted DIA processing algorithms are often not specific enough to differentiate between multiple versions of the same peptide differing only in the PTM site localization. Spectronaut's PTM localization workflow for DIA allows to benefit from the sensitivity, accuracy, and precision of a DIA targeted extraction with a high confidence site localization.

The PTM localization algorithm for peptide-centric analysis utilizes information not typically available to a classic DDA. This includes a usually full isotopic pattern for all fragment ions and the possibility of generating short elution chromatograms to correlate with the targeted peak shape. The later allows for systematic removal of any interfering fragment ions that one could not account for in DDA. Combining those two unique aspects with additional scores based on fragment mass accuracy and intensity shows excellent performance.

The novel PTM localization algorithm can work on any variable modification (such as phosphorylation, methylation, acetylation, sulfation, etc.) or combinations of different modifications and does not require specially generated libraries.

Reviewing your Analysis

The Analysis perspective features Grid and Tree Views (Figure 3). The Grid View is a protein centric view of the analysis that allows easy visualization of differential protein expression (Figure 19). Differential abundances across samples are displayed in colours: blue for lower and red for higher abundances. You can also filter the list by identified PGs, complete profiles, coefficient of variation, by the candidates according to the differential abundance testing. By dragging a column and dropping it to the designated area above the grid, you group the list by that function. For a selected PG you can also display protein coverage, quantity profile, and

condition box plot. Grid View data can be easily exported using a dedicated Grid View Report schema in the Report perspective.

Figure 19. The Grid View is a protein centric view of the analysis and shows differentially abundant proteins across conditions – low abundant in red and high abundant in blue (if compared against the mean for that protein group), protein coverage, and quantity profiles (log2 quantities).

The Tree View shows the data organized in an expandible tree (left panel) and corresponding plots, reports, and summaries (right panels) (Figure 20). By default, the hierarchy of the tree is:

>Run
  >Protein Group
    >Elution group
      >Precursor
        >Fragment ions

The runs are by default filtered by identification, which means that only what has passed all the identification thresholds specified in the settings is shown (see DIA identification settings in Appendix 7.1). These include: precursor posterior error probability (PEP) cutoff, precursor Q-value cutoff, and protein Q-value cutoff at both experiment and run level. You can also see what has not been identified (“Not Identified”), or remove all the identification filters and visualize all the data (“Not Filtered”).

You can change the tree structure by right-clicking on the experiment tab → Group by, then select one of the options (more about these functions in Appendix 6. Experiment Tab Options, Section 7.6). The most common actions are also accessible through intuitive icons that are available under the experiment tab (Figure 20).

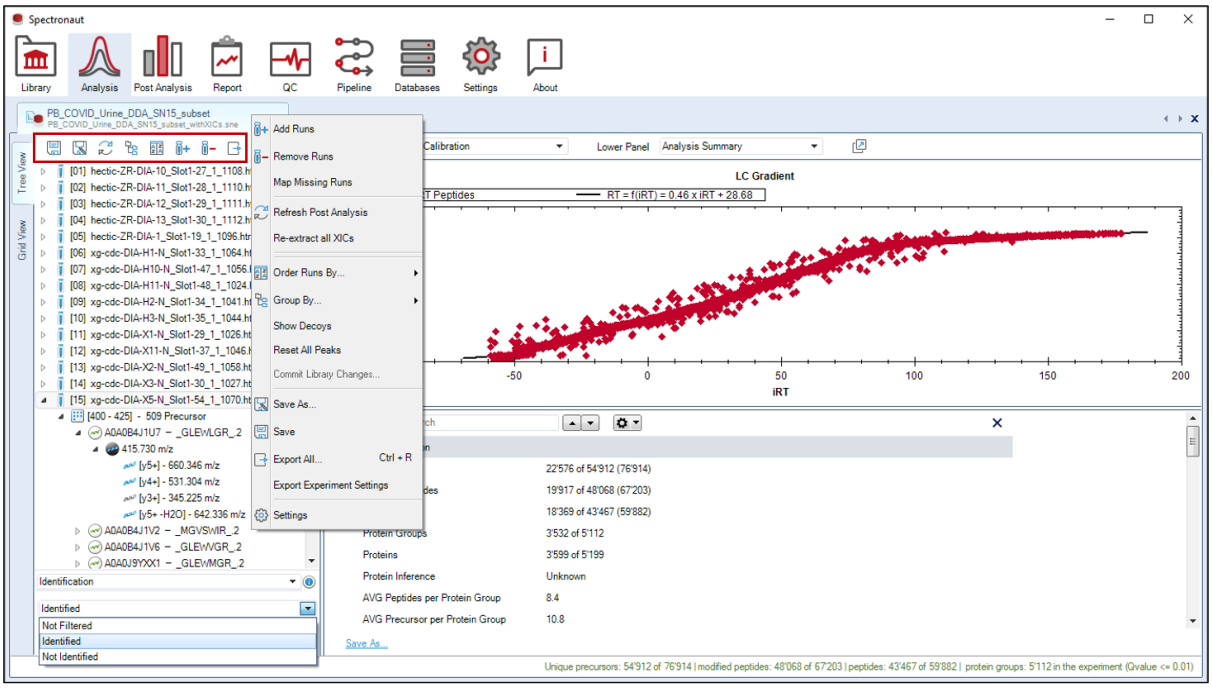


Figure 20. The analysis Tree View. The data tree is displayed on the left side. Plots and summaries are on the right side. The drop-down menu in the bottom-left corner allows to runs filtering. A summary of the number of identifications is shown in green at the bottom.

In the data tree, you can right-click on any of the elements and execute element-specific actions such as accepting or rejecting a precursor ion or refine the fragment ions selection (Figure 26).

The right side of the Tree View is divided into upper and lower panel to display two different plots at the same time. By selecting the same plot in both panels, you will get it in large view. The plots change based on the selected element in the data tree (e.g. run-level, protein-level, precursor ion level). To know which plots are available at each level, see Appendix 5. Analysis Perspective Plots (Section 7.5).

To visualize more than two plots simultaneously, you can use the floating plotting windows (Figure 22). You can open up to three floating widows per experiment.

Figure 21. Filtering the tree in the Analysis Perspective. Check the box for a filter and give the corresponding value. An example of filtering for a peptide sequence is shown on this figure.

Figure 22. Visualization of several plots simultaneously on one or many monitors with the floating plotting windows.

You can also visualize several perspectives simultaneously by detaching them (Figure 23). This allows, for instance, to keep the Analysis and Post-Analysis perspectives open on separate monitors. To detach a perspective, select it and press the F12 key on the keyboard.

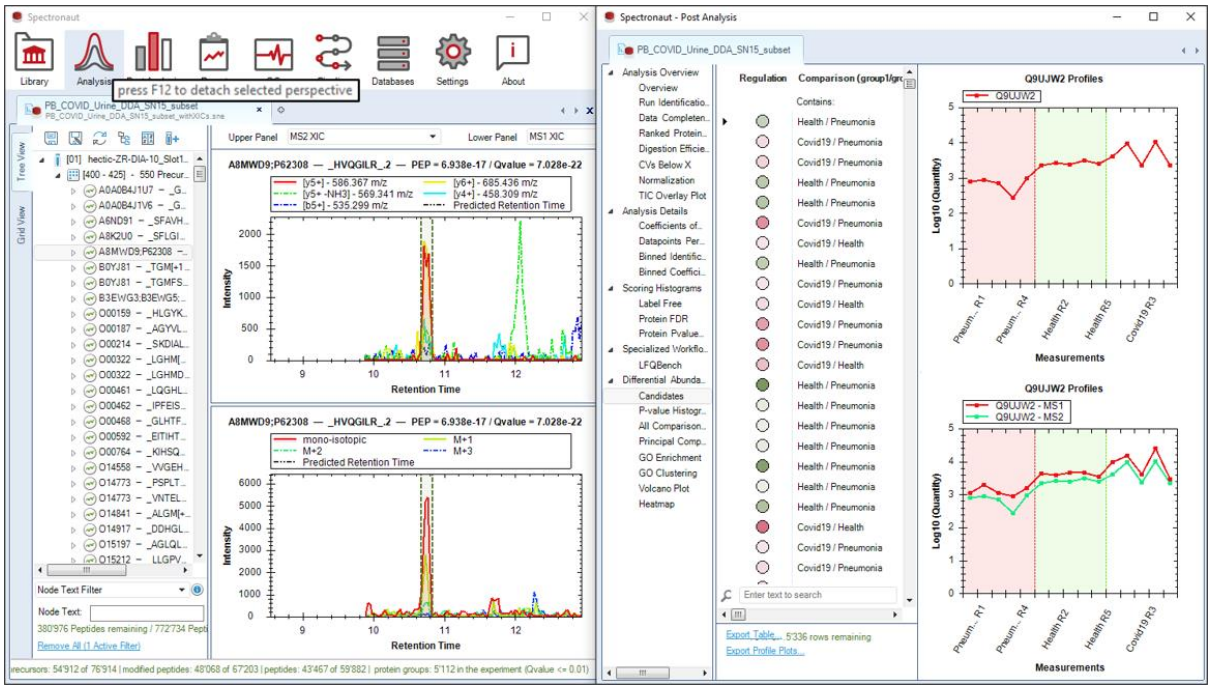


Figure 23. Perspectives can be detached and visualized simultaneously.

3.4.1.7 Analysis Perspective Plots

Spectronaut provides a comprehensive set of plots and reports for the review of the analysis at different levels: run, precursor and fragment ion.

- Run level plots: information about the calibration status, DIA method used, TICC, run meta information and cross run performance
- Precursor and fragment level plots: XIC chromatograms, score-centric plots and cross-run profile visualizations. The latter ones are only available in multi-run experiments and disabled for experiments containing only one run or peptides that are only targeted in one run.

Please, visit the Appendix 5. Analysis Perspective Plots (Section 7.5). to find an example and description of each plot.

To learn some tips about how to use the plots in Spectronaut, see Box 8.

3.4.1.8 Tree Filtering

One can apply one or several filters to the data tree. These filters only influence what is shown in the Analysis Perspective but not, for instance, the Post Analysis Perspective. Select a filter from the drop-down menu (Figure 21) and set the filter criteria. The filter is now checked within the drop-down menu. To combine filters, select a different filter and define the value that should be applied. A precursor must apply to all selected filters in order to be shown in the review tree. By default, the Identification filters apply (see Section 0).

Note: sometimes is not obvious that a filter is applied. Make sure you check the filter list before reviewing your analysis further.

You can specify a custom criterion with the "User Group" filter. This value can be set during the library import by selecting a specific column as "User Group".

3.4.1.9 Experiment Tab Options

Right-clicking on the experiment tab in the Analysis Perspective opens a context menu with many functionalities that can be applied to the experiment (Figure 20). The most common actions are also accessible through intuitive icons. To see the full details for these options, refer to Appendix 6. Experiment Tab Options (Section 7.6). Some of the most relevant ones are:

1. **Save and Save as:** Spectronaut will not save the analysis automatically. To save an analysis you will have to do it manually. You can save your analysis **with or without ion traces (XICs)** (Figure 24):
   a. ***With ion traces (FULL)****:* the file generated (*.sne file) will be larger, but Spectronaut will not require the run files to be available when you load your saved analysis again.
   b. ***Without ion traces (XICs)****:* The *.sne file will be smaller, and you can map the run files after loading it.

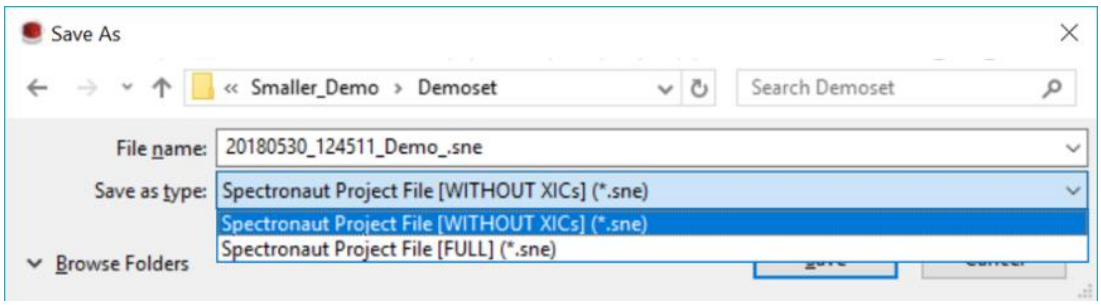


Figure 24. Saving your *.sne file with or without ion traces (XICs).

2. **Group by:** change the structure of the data tree. The main level will still be "Run". See details in Appendix 6. Experiment Tab Options (Section 7.6).
3. **Settings:** this option allows you to review and **change** many of the analysis settings without having to run the analysis again. For instance, you can change the FDR cutoffs, quantification settings, FASTA file for protein inference, conditions set-up, and many more (for details see Appendix 6. Experiment Tab Options, Section 7.6).

3.4.1.10 Manual Analysis Refinement

Spectronaut is implemented with a highly reliable peak picking algorithm and with scores and confidence thresholds, adapted to high throughput workflows with several 1000s of runs analyzed, several 1000s of protein identified and 100000s of precursors targeted. Reviewing the results of your analysis is in general not expected or needed.

However, Spectronaut provides the possibility of exploring your results and fine tuning several aspects manually, if you would like to do so (see Box 11 on how to optimize the manual reviewing process). For this purpose, the Analysis Perspective provides a number of aspects you can interact with and modify. The most relevant are:

- **Refine elution group integration boundaries.** Select an elution group in the data tree and set the right-side plots to, for example, MS2 XIC. You will see the ion chromatogram corresponding to the selected elution group within two green lines (Figure 25). You can manually slide these lines on both sides of the integrated peak, to set different boundaries. A new q-value should be calculated. A hand icon will appear next to the elution group in the data tree, denoting it was manually modified.

Figure 25. Reviewing Spectronaut peak picking. Integration boundaries can be modified by dragging them. The precursor will be marked as manually modified.

- **Manually select a different peak in the XIC**. Select an elution group in the data tree and set the right-side plots to, for example, MS2. Hover over the peak you want to assign and click when the cursor changes into a hand. The integration will be transferred, and a new q-value should be calculated. Similar to the action above, the precursor will be marked as manually modified in the data tree.
- **Manually accept or reject an elution group:** If you right-click on an elution group, you can manually accept or reject it. The icon next to the precursor will change to denote it has been manually modified (Figure 26).
- **Manually define an interfering fragment ion:** in the data tree, when you expand the precursor, you can see the fragments present in the library for that specific precursor. The ones used for quantification will have a blue icon, while the ones detected as interferences have a grey icon (Figure 27). You can define interferences manually by right-clicking on the fragment ion and unchecking the "Used for Quantification" option.

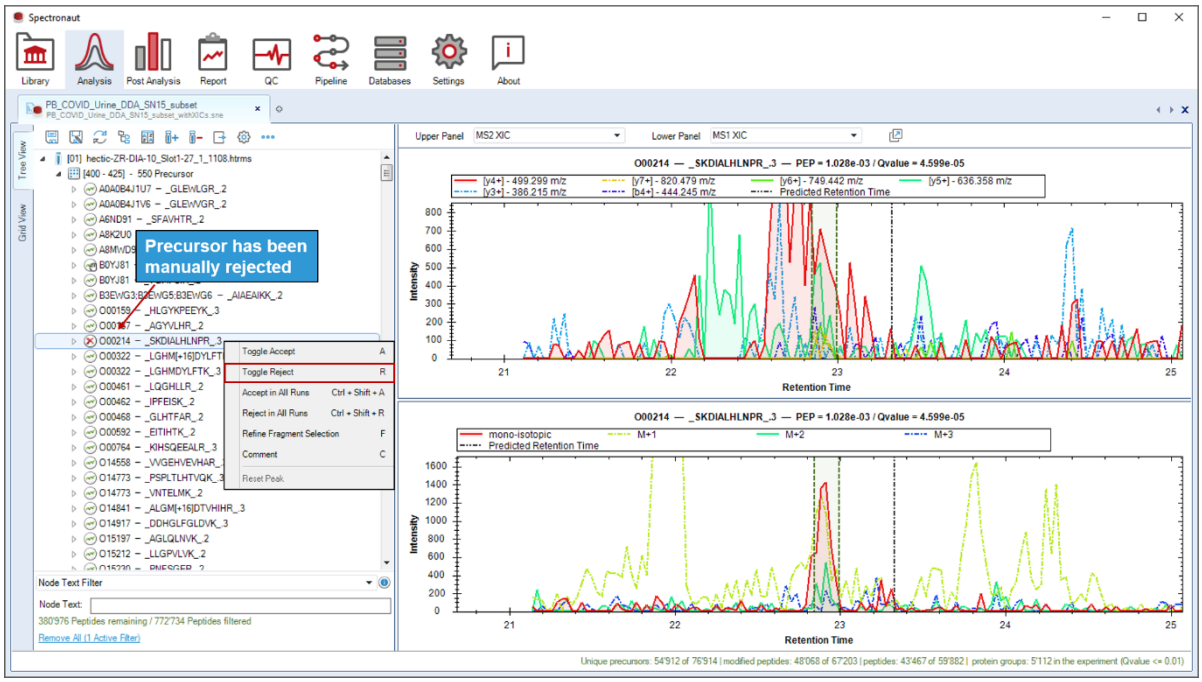


Figure 26. Reviewing Spectronaut peak picking. Precursors can be manually accepted or rejected by right-clicking and choosing the option. The precursor will be marked as manually accepted or rejected.

Box 11. Tips to optimize manual reviewing of your data (UI responsiveness)

If you need to manually review and actively navigate through your analysis in the Analysis Perspective, you might find some processes to be a bit slow and the software not as responsive as hoped. There are several things you can check in order to make the process as fast as possible:

1. Have your run files locally. Having your run files on a network drive is not recommended and can significantly slow down the computational processes.
2. Convert the run files to HTRMS files (see Section 4) before running your analysis.
3. If you saved your *.sne file without XICs, re-extract your XICs (see Section 3.4.1.9).
4. Group your data tree by precursor window.

Happy reviewing!

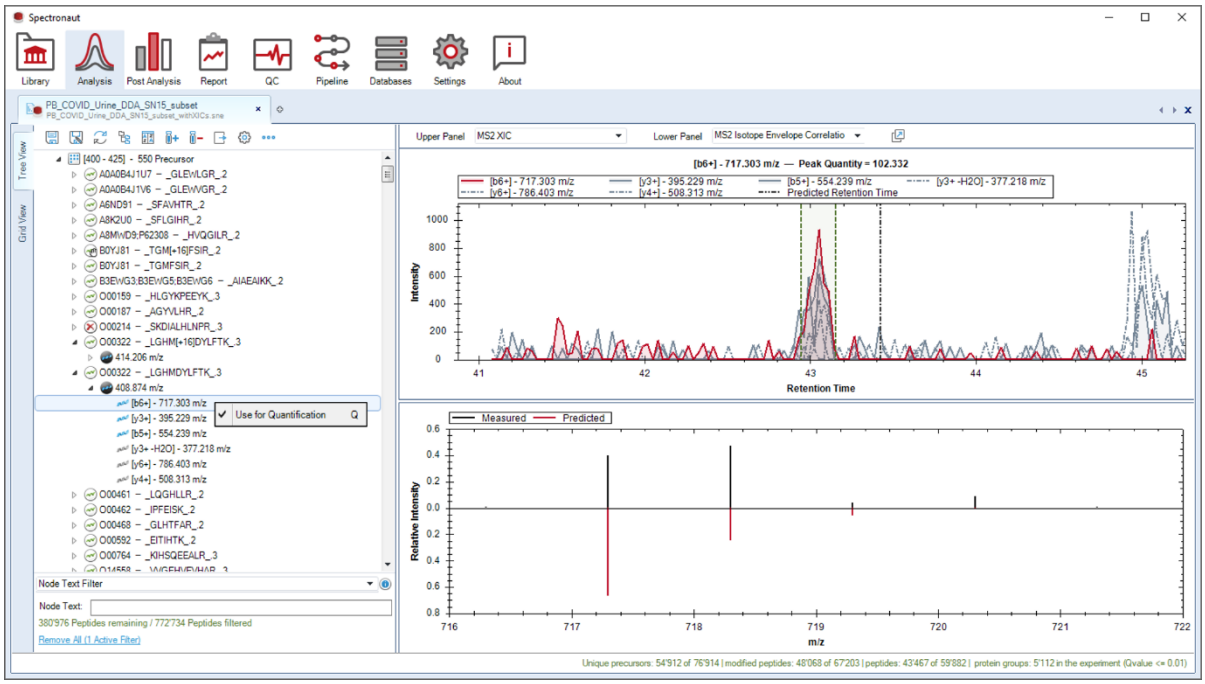


Figure 27. Manually define interfering fragment ions or manually accept fragment ions for quantitation that were defined as interferences by Spectronaut.

3.4.1.11 Library Refinement

Spectronaut allows you to refine the fragment ion selection of your libraries. You can remove fragments that show interferences, specifically select fragments to cover interesting modification sites, or add additional fragments that were not detected in the original DDA analysis but are nicely visible in DIA.

To perform library refinement, the peptide assay must originate from a library generated in the Library Perspective. Right-click on the elution group node and select "Refine Fragment Selection". A new dialog will appear that shows the selected peak in detail (Figure 28). The list on the left will show you all the fragments of this peptide that are present in the library, and which of those are currently selected. Additionally, a list of theoretical fragments is generated. In order to change the selection of theoretical fragments, right-click on the "Theoretical Fragments" node and select "Set Fragment Filter". This will allow you to expand the set of theoretical fragments so that it will contain different ion types, as well as common loss types.

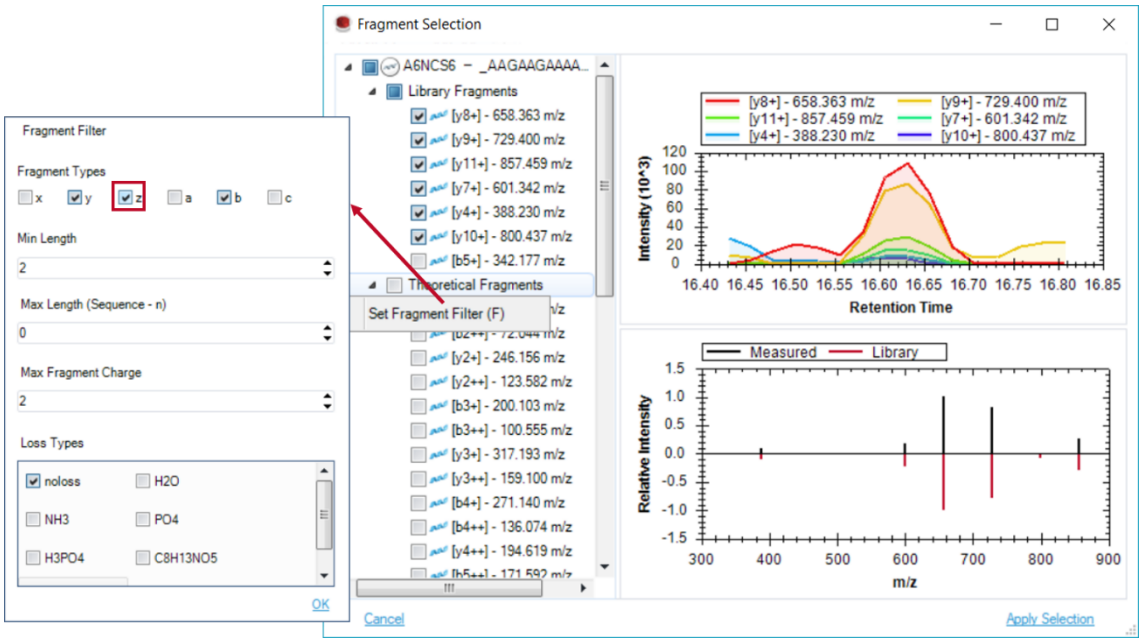


Figure 28. Refine your library while reviewing your analysis. You can, for example, use different ion series (z-ions) by adding them into the fragment tree and selecting them. A preview of how the XIC looks with the current selected ions is shown on the right. You can also look at how individual ions XICs appear by selecting individual ions.

On the right, you can see a preview of how the fragment selection will affect both the XIC and the match between the predicted and measured fragmentation pattern. Please note that theoretical fragments will not contain a predicted intensity. To add or remove a fragment from the spectral library, simply check or uncheck the corresponding fragments.

After the new fragment selection is complete, click "Apply Selection" in the bottom-right corner. Please note that the refinement is not performed immediately. In order to effectively change the selection in your experiment, as well as in the spectral library, right-click on the experiment tab in the Analysis Perspective and click "Commit Library Changes…".

This extra step is necessary since the update will require a re-extraction of the affected peptide from all currently loaded runs. Spectronaut will, therefore, perform this operation in batches once all manual fragment selection is done, and not for each peptide individually. All changes made to a peptide's fragment selection will only take effect once the library changes are committed.

After clicking "Commit Library Changes…" another popup window will appear asking you for a version name. Spectronaut features a version control for spectral libraries that allows you to switch between different versions of a given library. This way, any changes to the original library can be reverted. In order to change the version of a library, go to the Library Perspective, right-click on the respective library node and select "Set Selected

Version". A small window featuring a drop-down list will allow you to select which version of this library to use.

3.5 Post Analysis Perspective

The Post Analysis Perspective in Spectronaut® reports processed results for your analysis. It shows summary information about identification, quantification, and results from the differential abundance test, hierarchical clustering, principal component analysis and GO terms enrichment and clustering (Figure 29). Moreover, when the PTM analysis in the PTM workflow settings is selected, Spectronaut provides differential abundance results at the PTM site level. For example, the PTM vs protein fold changes plot, showing protein group log2 ratios plotted against log2 ratios of modification sites, helps to identify changes on the modification site level that are independent of changes of the protein abundance.

Furthermore, the digestion efficiency plot, especially useful for LiP workflows, shows the number of peptides per digest type and enzyme. Finally, under specialized workflows, a plot dedicated to LFQ benchmark studies is available (Figure 87).

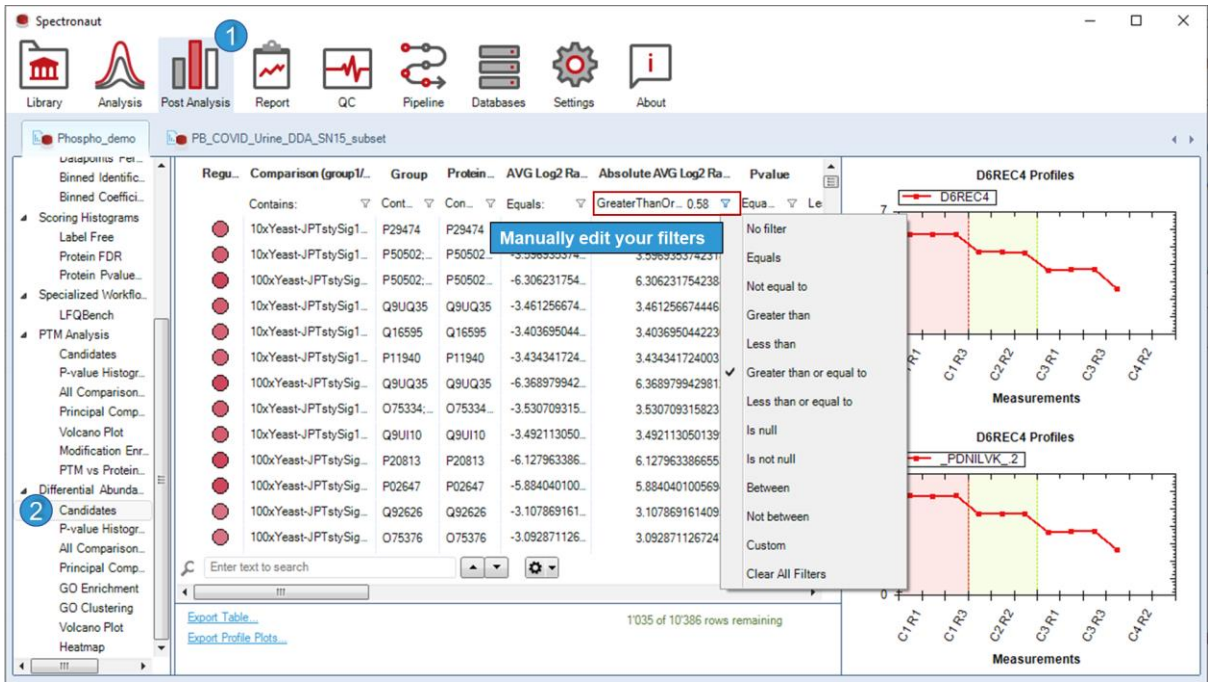


Figure 29. Post Analysis Perspective. Several summaries, tables, and plots are available as you navigate through the nodes in the tree. In the figure, you can visualize the Candidates list. You can modify your candidate set by filtering directly on this table. The default applied filters are a 1.5-fold change and a q-value of 0.05.

Analysis Overview

Here you will find some experiment-wide information that will give you a rough idea about dataset characteristics. Under Overview, you will see a summary with the number of proteins and peptides identified (by conditions), number of proteins and peptides quantified (according

to chosen quantification settings), miss-cleavages, library recovery, and other general metrics about the experimental outcome. To support this overview, you will find several plots related to the number of identifications, the data completeness, the coefficient of variation, and the normalization. On each of these plots, you can change many settings by using the right-mouse click option. To see the full details of each plot, see Appendix 7. Post Analysis Perspective Plots (Section 7.7). Learn more about plots in Spectronaut in Box 8.

Scoring Histograms

Under this node, you will find plots related to the behavior of the target and the decoy distribution estimation. This behavior defines the discriminant scores (Cscore), q-values (Qvalue), and sensitivity on the precursor level. Scoring histograms are shown for each workflow in the current experiment depending on your experimental set-up. Find all the details about these plots in Appendix 7. Post Analysis Perspective Plots (Section 7.7).

Analysis Details

The Binned Identification plots in the Analysis Details shows the relative number of precursor identifications across conditions binned according to three variables: iRT, intensity and m/z. Relative ID means the portion of measured precursors that were confidently identified in a bucket i.e. passed the identification cutoffs defined in the experiment settings. These plots sprovide valuable feedback on the performance of the measurements for different conditions according to technical criteria such as liquid chromatography and mass spectrometer performance. Datapoints Per Peak plot shows data points per peak in each condition as well as their distribution with the median (visualized in box plot or violin plot). The Binned Coefficients of Variation show CVs within conditions and are binned according to the same three variables: iRT, intensity and m/z. See these plots in Appendix 7. Post Analysis Perspective Plots (Section 7.7).

Differential Abundance

3.5.1.1 Candidates table

The results of the differential abundance testing will show up under this node. The Candidates node shows a table with the results, annotated by paired or unpaired t-test comparison (Figure 29):

- The direction and the percentage of change are noted by color and color intensity, respectively; the level of significance is noted by the size of the circle.

- The fold changes are expressed as log2 transformed ratios of averaged replicates (AVG Log2 Ratio).
- The table is, by default, filtered by a q-value (multiple testing corrected p-value) of 0.05 and an absolute log2 ratio of 0.58. You can change these filters to your preferred cutoffs. The filters applied to this two metrics in the table will automatically apply to the volcano plot as well.
- You can add and hide columns in this table by right-clicking on any of the headers and selecting Column Chooser. For example, you may want to add the p-value column.
- It is possible to search any character in the table with the Search field at the bottom of the table.

The candidates table can be exported as an excel file by clicking on "Export Table…" at the bottom.

In addition to the table, the candidates will be shown as plots on the right side. These plots can be customized in several ways by right-clicking on them and choosing your preferred options.

3.5.1.2 Principal Component Analysis

Principal Component Analysis plot shows clustering of the samples based on their protein profiles.

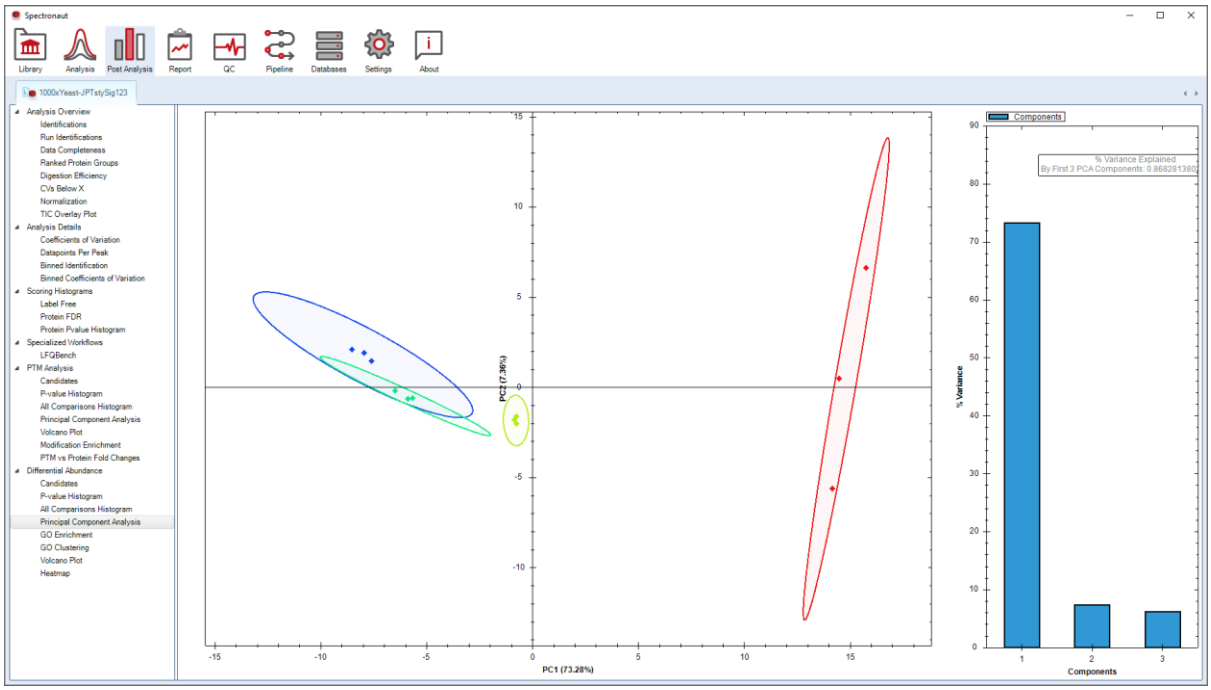


Figure 30 The Principal Component Analysis plot shows clustering of the samples based on their protein profiles. The components bar plot on the right side shows how the first three

components explain the variance. By right clicking on the plot you can perform several actions, such as Save Image As, export the data matrix, or modify the scaling.

3.5.1.3 GO Enrichment

Under the Differential Abundance node, you will also find the results from the Gene Ontology (GO) term enrichment and the GO term clustering. If you added GO annotation to your experiment, either within the library or during the analysis set-up, Spectronaut will perform a GO term enrichment test.

Spectronaut comes with the human GO annotation implemented. If you are working with a different organism, you can download your relevant annotation from http://geneontology.org/page/download-annotations or any other source, and import it into Spectronaut via the Databases Perspective (Databases Perspective → GO Databases → Import Gene Annotation). Your annotation is now ready to be appended to a library during the library generation, or to be selected during the analysis set-up. Please note, that GO annotation file you will use for the library and for the analysis set up should match the FASTA file selected for those steps.

The term enrichment test will check whether there are biological processes, functions or cell compartments over or under-represented within the candidate set. In other words, it will highlight processes, functions or compartments affected by the experimental conditions.

During the analysis, the first step is to determine how frequently a GO term occurs in the background proteome, i.e., all proteins identified throughout the whole experiment. Based on this information, this term is expected to be found a certain number of times in a random set of a given size. If the GO term occurs more frequently in your candidate set than expected, it is considered as significantly overrepresented; if it occurs less frequently, the term is considered as significantly underrepresented (Mi et al., 2013). The level of significance is given by a p-value. Spectronaut will perform two multiple testing correction methods to this test: Bonferroni (Dunn, 1961) and Benjamini-Hochberg (Benjamini & Hochberg, 1995), for which the corresponding corrected p-values are also displayed.

If you change the candidate set, i.e., you apply a different filter in the Candidates table, the enrichment must be recalculated.

The result of the enrichment test will be shown as an interactive table where you can group the results according to a column or filter according to any of the features (Figure 31). Similar to the Candidates table, you can easily search within the table with the search field at the bottom. The GO Enrichment table can be exported by clicking “Export Table…" below the table panel.

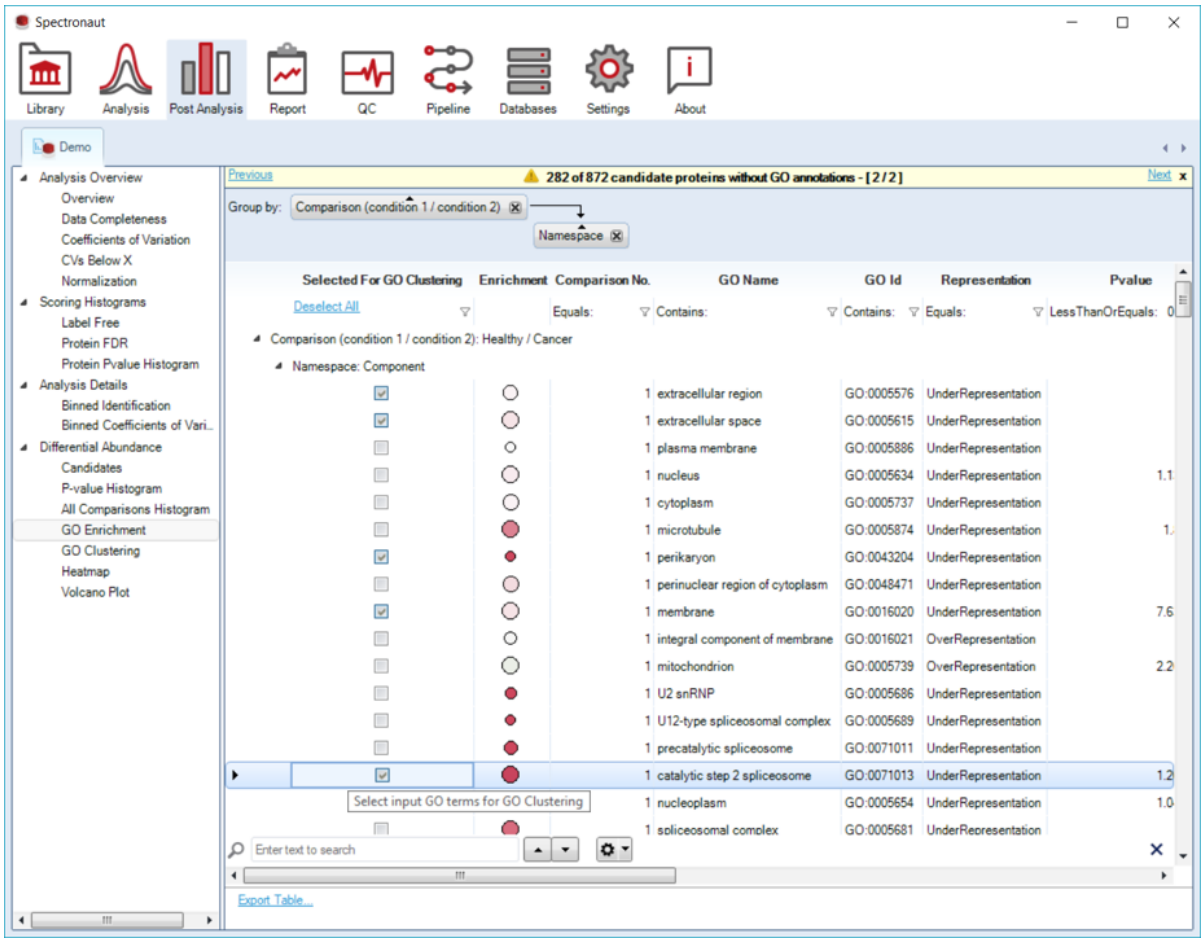


Figure 31. GO term enrichment result. Similar to the Candidates table, each column can be filtered according to several options. You can also group the results by any of the column headers by dragging them to the Group by field. If you want to do GO clustering on manually selected terms, use the first column of this table.

3.5.1.4 GO Clustering

GO clustering is a step further towards reducing the complexity of the differential abundance test results into an easier to interpret picture. If your GO term enrichment seems too convoluted, GO clustering will group related terms by similarity. The result is a shorter list showing groups of GO terms. GO clustering works based on the REVIGO algorithm (Supek et al., 2011). The semantic similarity of two terms is calculated based on their position/relation in the Gene Ontology graph.

Spectronaut will perform a GO term clustering on a subset of the terms from the enrichment analysis. This subset can be defined in two ways:

1. Manually selecting them in the GO Enrichment node, by activating the check box in the first column (Figure 31).
2. Filtering from the GO enrichment node by:
    - Namespace: biological process, molecular function, subcellular compartment
    - Type of representation: over, under, or both
    - Number of terms to cluster, ranked by p-value
    - Fold change
    - Number of proteins per term

In both cases, you will have to specify:

- The paired or unpaired comparison you want to look at
- The term similarity cutoff you want to apply (being 0 lowest and 1 highest)
- Whether you want to apply a p-value cutoff and which one (corrected or not): If two terms are clustered, one of the terms will be chosen as cluster representative. The other will be discarded. Usually, the more specialized term "wins" since it carries more meaning (such as ER lumen over cell). However, if you use p-values, the term with the lower p-value will win, even if it is a very general term.

When you are ready, click on "Start Clustering…"

The results show a list of GO terms (cluster representatives) on the left, and the dispensed terms that were clustered underneath them. The Dispensability score shows at what similarity cutoff the GO term would be clustered under another term.

As usual, you can export the table of results as an excel sheet by clicking on "Export Table…" at the bottom.

3.5.1.5 Differential Abundance Plots

Under the Differential Abundance node, several plots related to the post-analysis are also generated. Please, find detailed information on each of these plots in Appendix 7. Post Analysis Perspective Plots (Section 7.7). The most relevant are the Heatmap and the Volcano Plot:

1. The Heatmap will be clustered row and column wise according to the Post Analysis settings. The raw data of the Heatmap can be exported via right-click on the plot (Figure 32)

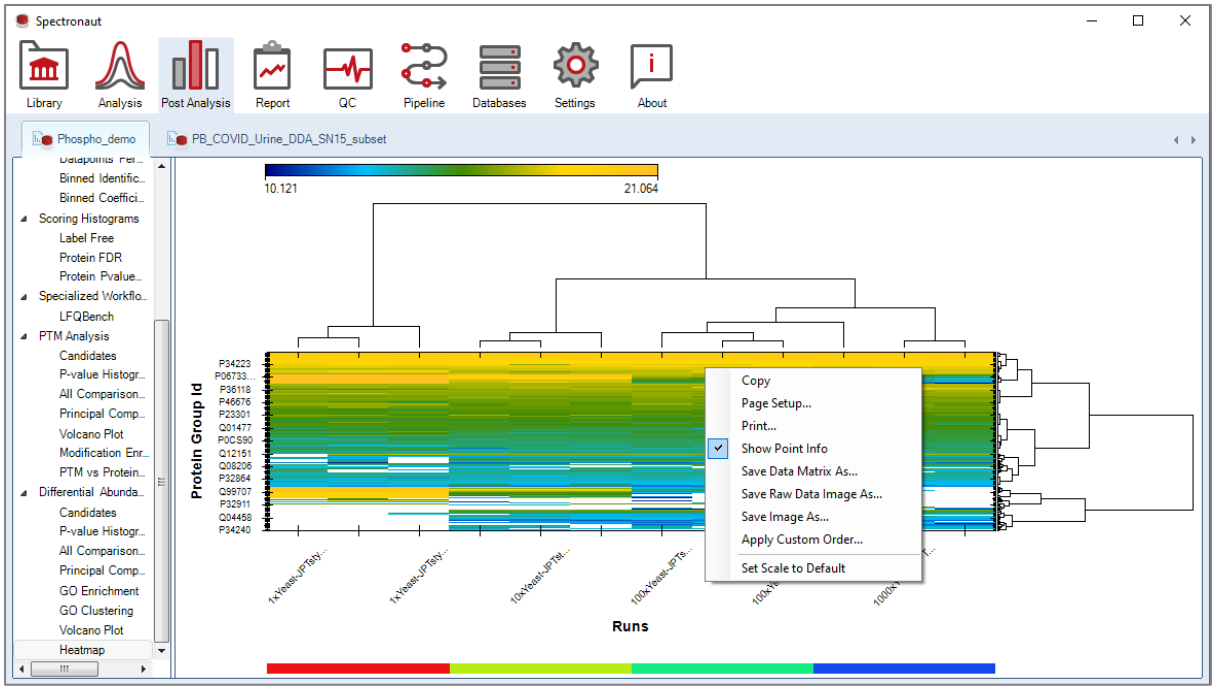


Figure 32. Heatmap with clustering in both rows and columns. The heatmap is built using the set of confidently identified datapoints. By right-clicking on the plot you can perform several actions, such as Save Image As, export the data matrix, or modify the scaling.

2. The Volcano Plot shows the results of the differential abundance test by plotting the peptides or proteins' fold changes against the significance level. The candidates will appear in red on the plot (Figure 33). By selecting one of the boxes above the plot, you can display annotations for all differential analysis candidates, or you can custom select proteins of interest and highlight them in blue.

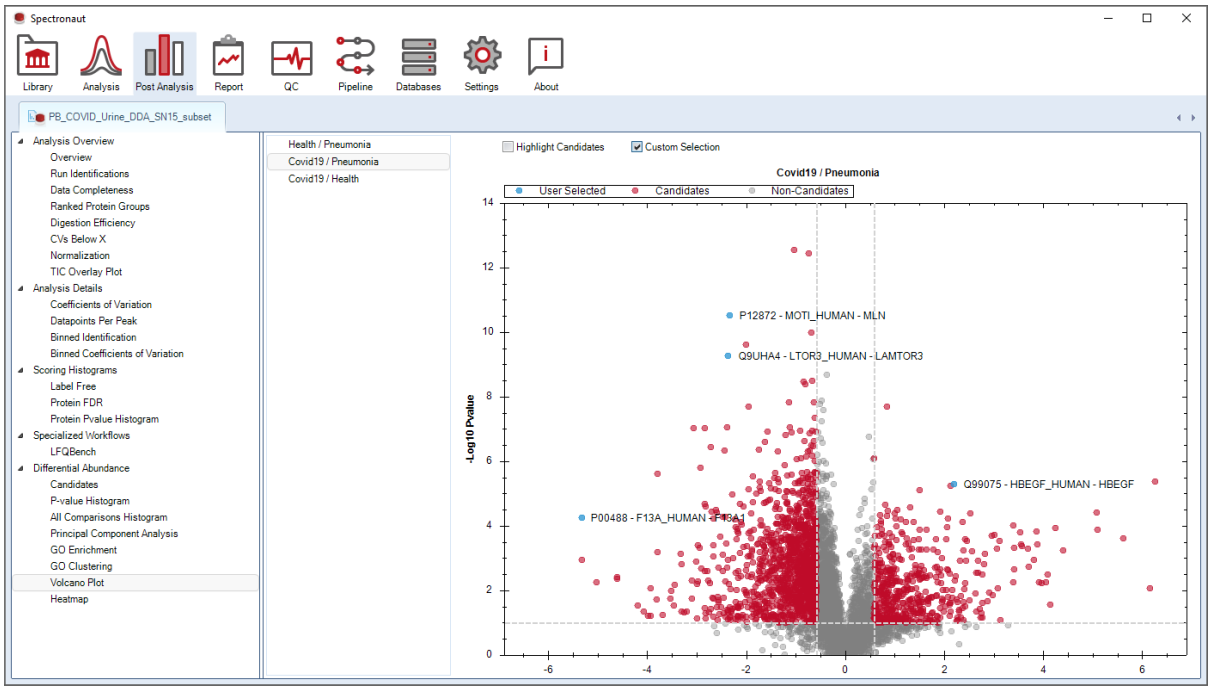


Figure 33. The Volcano Plot shows the candidates in red. This plot is updated when you modify the Candidates table's thresholds. By right-clicking you can choose several actions, such as deactivate the legend or change the scale in the graph.

PTM analysis

The PTM analysis results are available for the experiments where the PTM workflow settings were chosen. The new Spectronaut 15 performs PTM differential abundance analysis on a modification site level (for more details see section 3.4.1.6)

The results of the PTM differential abundance analysis are reported in the candidates list and corresponding volcano plot. Among other information, the PTM analysis node contains principal component analysis that shows clustering of the samples based on their PTM sites quantification profiles. Additionally, PTM analysis specific graphs are available: Modification Enrichment (see section 3.5.1.7) and PTM vs Protein Fold Changes (see section 3.5.1.8).

3.5.1.6 Candidates table

The results of the differential abundance testing on a modification site level will show up under this node. The Candidates table contains a list of differentialy abundant modification sites with their fold changes and Qvalues, annotated by paired or unpaired t-test comparison. The identification key of the specific modification site object, for which differential abundance analysis result is reported, could be found in the column «Group».

The PTM analysis candidates table can be viewed, modified and exported in a similar way as candidates of differential abundance analysis on the protein level (see section 3.5.1.1). The candidates of the PTM differential analysis will be visualized in the corresonding volcano plot which example is available in Appendix 7.7.

3.5.1.7 Modification Enrichment

The plot shows the percentage of all identified precursors that are carrying a selected modification in each of the experimental runs. If the modification can occur on different amino acids, the plot will show percentage of the precursors carrying that particular modification on each of those amino acids. The example of such a graph is presented in Figure 34, showing an enrichment of phosphorylation localized on tyrosine, serine and threonine. The modification enrichment plot is dedicated to experiments conducted with the step of modified peptides enrichment.

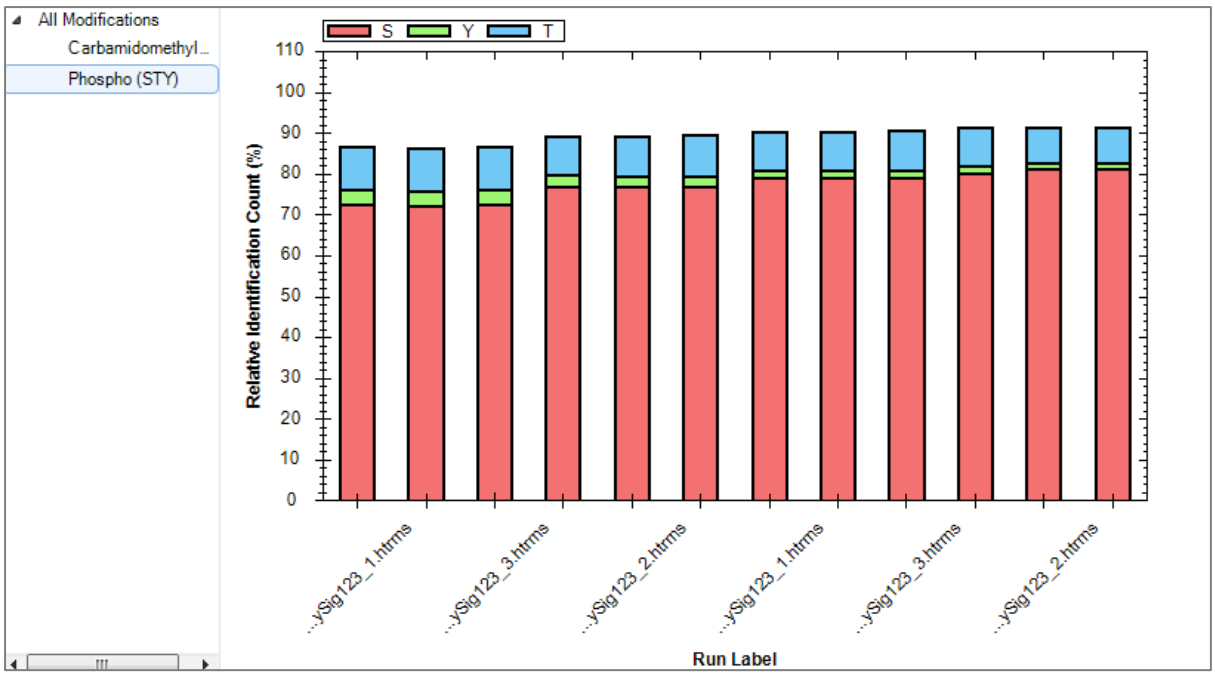


Figure 34. Modification Enrichment plot shows the percentage of all the precursors carrying given modification in each of the experimental runs.

3.5.1.8 PTM vs Protein Fold Changes

PTM vs Protein Fold Changes plot shows protein group log2 ratios plotted against log2 ratios of modification sites. The plot will help to identify changes on the PTM site level that are independent of the protein group abundance changes.

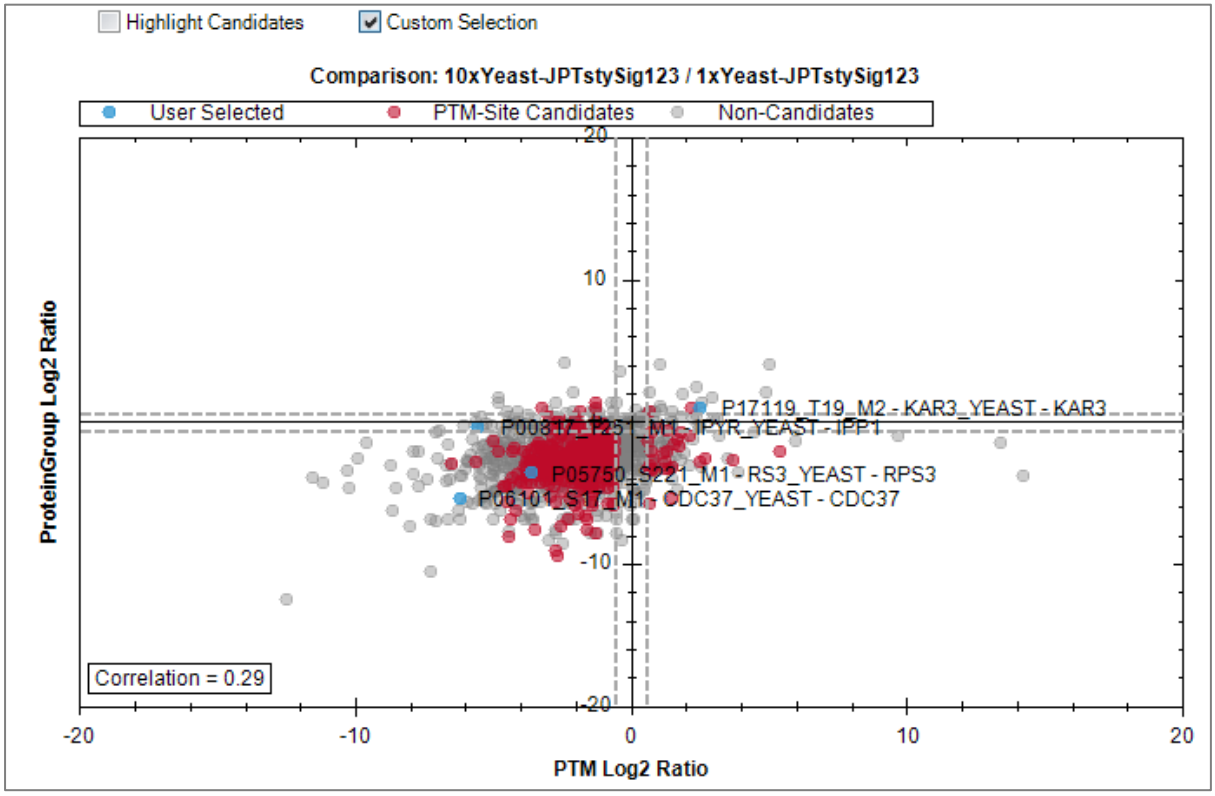


Figure 35. PTM vs Protein Fold Changes plot shows log2 ratio of the protein groups ploted against Log2 ratios of the PTM sites.

3.6 Report Perspective

Spectronaut® has a very powerful reporting strategy. In the Report Perspective, you can design and customize your report to contain any information you may need about the analysis. In case PTM workflow was used for the analysis, a specialized PTM site report will be available. Also, Grid View data can be easily exported using a dedicated Grid View Report schema. Report schemas of any report type can be saved and reused. You can also change the column names to fit your needs (Figure 36).

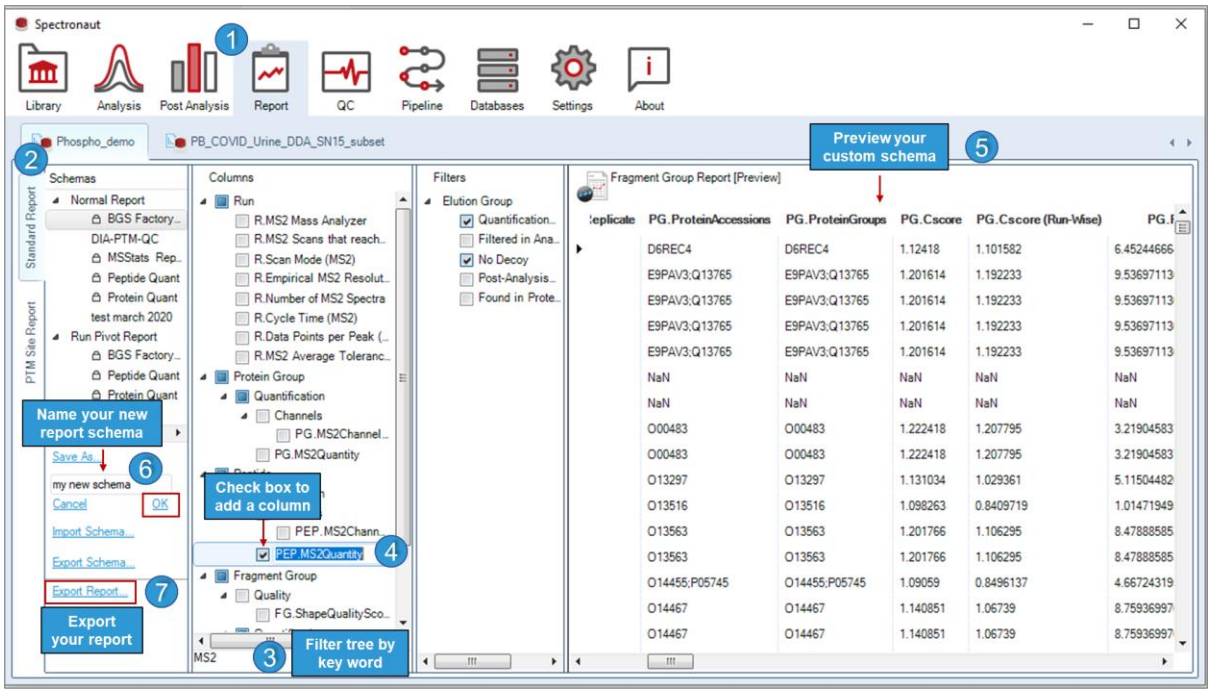


Figure 36. Report Perspective. The figure shows the process of customizing a Normal Report schema and exporting the data. Detailed explanation of the headers can be found by hovering over them or in Appendix 8. Most Relevant Report Headers

The Report Perspective contains four panels, from left to right:

1. The Schema tree: all different report building schemas. If you save a custom one, it will appear here.
2. Column chooser: all possible reportable elements with check boxes to add or remove them. Below this panel, there is a search field to help you navigate through the different fields.
3. Filters applied to the report.
4. Report preview: A preview of how your report will look like. This is very useful when you are modifying a schema. When you are happy with your report structure, you can export it by clicking on "Export Report…" in the bottom left corner to be able to see the whole matrix.

Report Schemas

Spectronaut includes several preconfigured reporting schemas that may fit most frequent needs. If you want to design your own, you can use one of the included as a base to build your preferred report.

Within the report schemas, there are two main formats you can export your data into: Normal Report and Run Pivot Report. Find detailed information about each format below.

Normal Report

In a Normal Report (long format), you will find each reported event in a single row (Figure 36). A Normal Report will usually have many more rows than a Run Pivot Report. This format is the one allowing for the most comprehensive report of your data. To build your Normal Report, add or remove columns from the Columns panel by checking or unchecking them (Figure 36). The Columns are organized by levels, from more general (Experiment) to more specific (Fragment):

>Experiment
>Run
>Protein Group
>Peptide
>Elution Group
>Fragment Group
>Fragment

Within each of these levels, the columns are again organized by categories (e.g., identification, quantification, scoring, etc.). The whole Columns tree is quite comprehensive, and expanding/collapsing categories when looking for a column can be cumbersome: to make the search for columns easier, there is a search field at the bottom of the Columns panel where you can type what you are looking for (Figure 36). Finally, to know which information a header contains, hover over it and you will see a text box popping up with a description.

To see a detailed description of some of the most relevant columns, see Appendix 8. Most Relevant Report Headers

Run Pivot Report

In a Run Pivot Report (wide format), each run (sample) will be a header column. You can choose which element you want to be rows in the Columns panel under Row Labels (e.g.,

stripped peptide sequence) and which value you want in the cells under Cell Values (e.g., quantitative value, Figure 37). If you choose more than one Row Label or Cell Value, the table will multiply its length column-wise. This report will probably have fewer rows than a Normal Report.

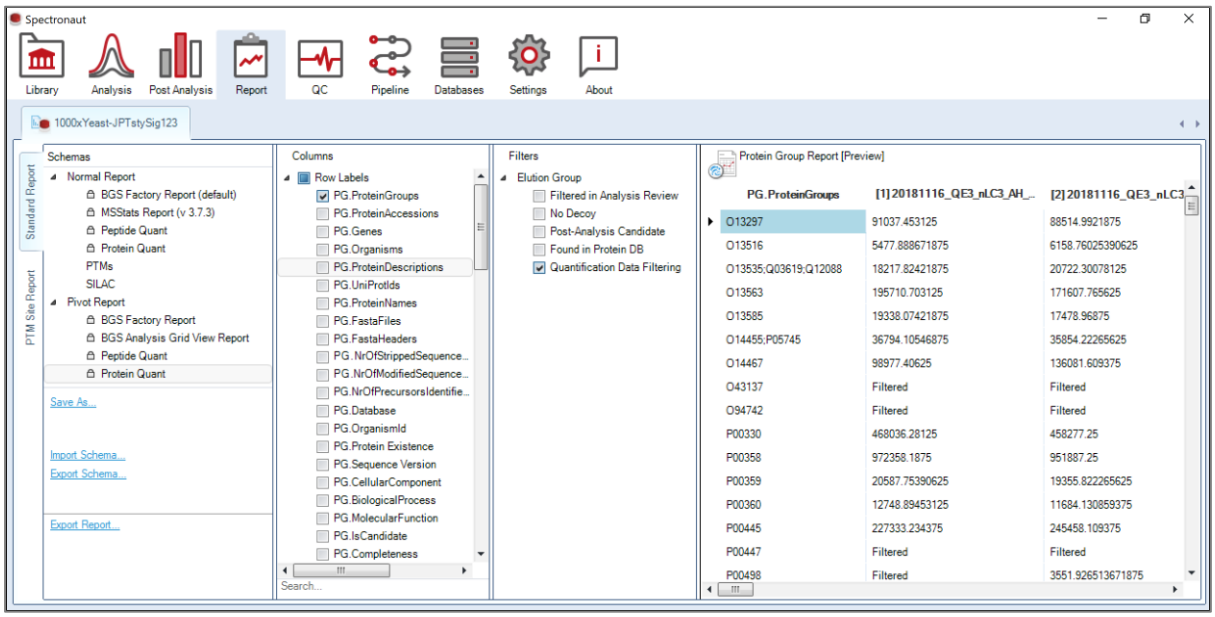


Figure 37. Run Pivot Report. This report is in wide format and contains one column per run (sample). The Grid View data can also be exported as a pivot report.

PTM site report

For the experiments analyzed with the PTM workflow, a specialized, PTM site report is dedicated. The Columns are organized by levels, similarly like standard report, from more general (Experiment) to more specific (PTM site):

>Experiment

>Run

>Protein Group

>PTM site

Among other information, PTM site level of the report contains details on how the precursor collapse was performed in order to obtain quantitative data for each PTM site object, PTM site quantitation data, PTM flanking region and PTM localization site probability. In order to obtain detailed description of some of the most relevant columns of PTM site report, see Appendix 7.8)

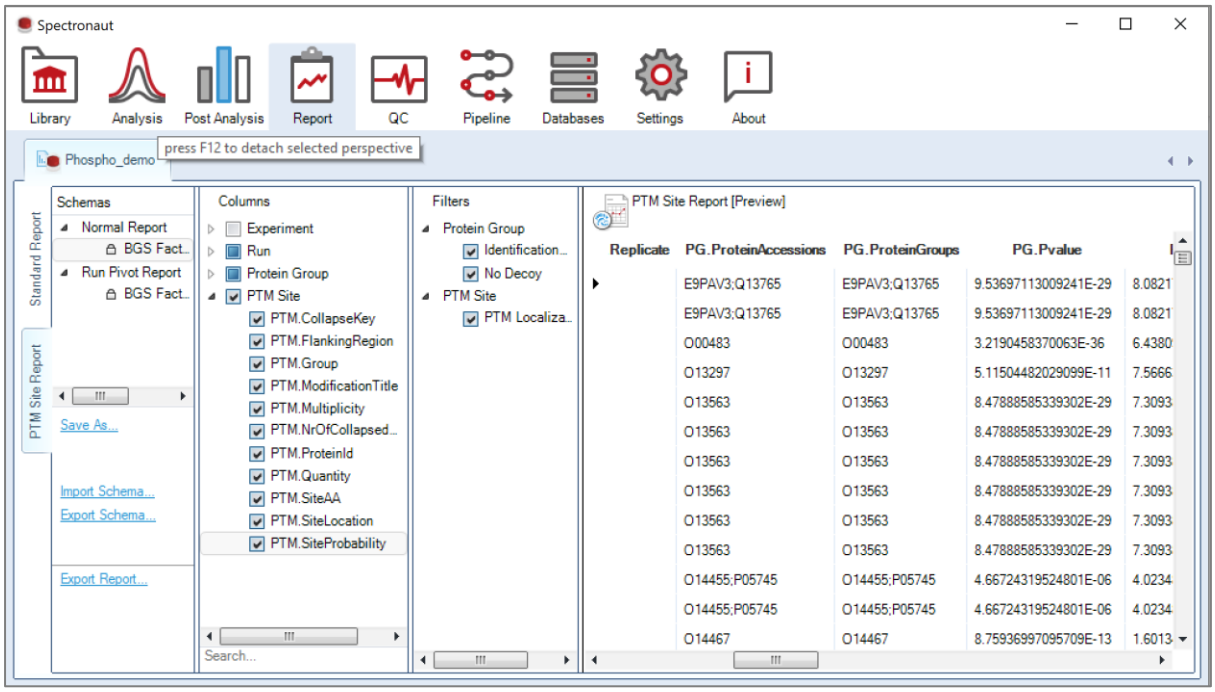


Figure 38. PTM site report. The report is avaiable for the experiments analyzed with PTM workflow.

PTM site report could be generated in format of normal or pivot report.

3.7 Quality Control Perspective

The quality control perspective of Spectronaut® is based on the iRT Kit Chromatography, mass spectrometer performance and analysis can be monitored over time using several performance indicators. Every successful analysis is stored in the quality control perspective (Figure 39). Spectronaut automatically detects various instruments and will create a separate quality control history for each of them. If you have more than one instrument of the same type it might be useful to rename them manually. Additional folder structures can be made according to the established quality control in a specific laboratory.

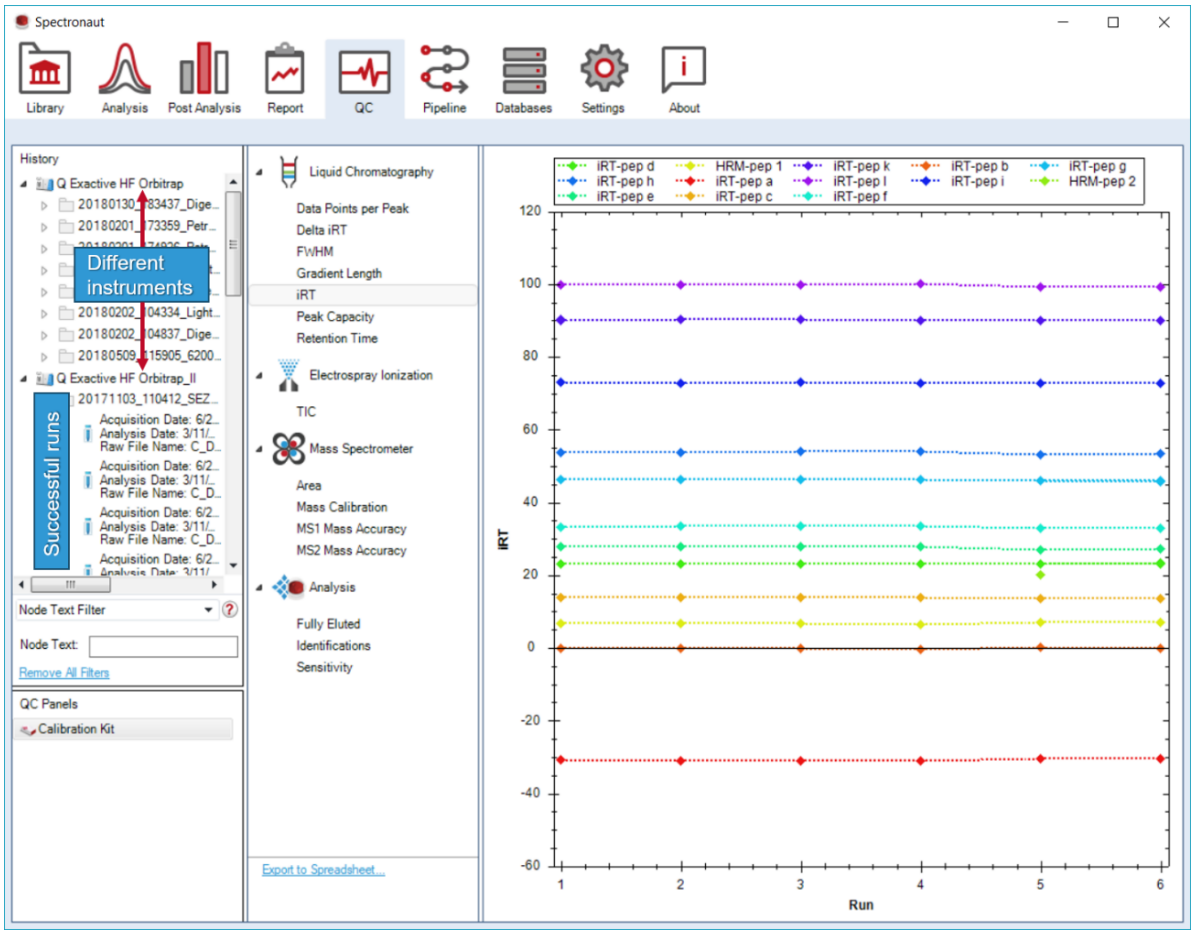


Figure 39. QC Perspective. Runs in which the QC panel is detected are saved in the History tree. You can monitor instrument performance with help of many plots related to several aspects of the experiment, from LC-MS to Spectronaut analysis.

Only as many runs as specified in Settings → Global → General → QC Plot History Length are shown in the plots.

QC Panels

In addition to the iRT peptides panel, sample specific QC panels can be created in the Library Perspective. Right-click on a spectral library and select "Enable QC". Whenever this

spectral library is used in an analysis, a respective QC file is written for each of the runs included in this analysis. When selecting this novel QC panel in the QC perspective, all the corresponding QC files (runs analyzed with the library enable for QC) will be available for QC monitoring

3.8 Pipeline Perspective

The Pipeline Perspective is used to batch process library-based DIA analyses using predefined settings. Spectronaut® works most efficiently when several experiments are processed sequentially rather than in parallel (because of disc IO). If you are not interested in manual evaluation of your peaks, the pipeline perspective might be your preferred choice. The set-up of an experiment works similarly to the set-up in the Analysis Perspective. The setup Analyses will be added to the Pipeline Queue. Clicking "Run Pipeline" will start to process the queued experiments sequentially. Spectronaut will automatically generate the report according to the settings in the chosen schema (Figure 40).

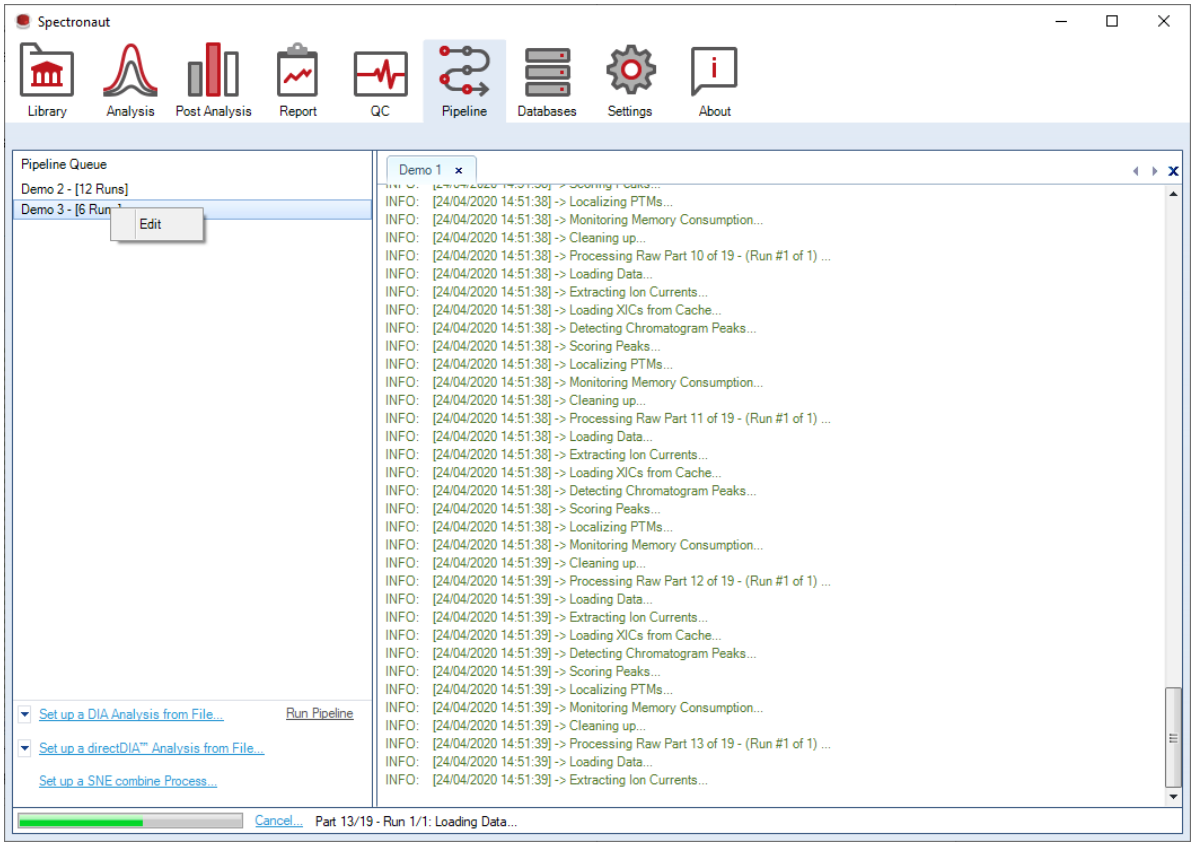


Figure 40. Pipeline Perspective. Queue DIA analyses so Spectronaut can process them sequentially. Experiment files, reports, plots, and summaries will be generated and stored according to the DIA Analysis Reporting Settings.

SNE Combine Workflow

Introduced with Spectronaut 14, the SNE combine workflow allows for a novel type of analysis workflow for processing large scale experiments. The idea behind this workflow is that a large experiment can be analyzed in small batches and the results stored as individual SNE files. You can then use the SNE combine workflow to merge the identification results of the individual batches in FDR controlled manner.

Additionally, SNE combine will run the same post processing steps as for a single batch experiment. These include cross run normalization, interference correction, protein inference, quantification, and FDR.

The entire process is geared towards memory scalability that will allow the analysis of 10'000+ DIA files on a low-cost desktop workstation.

In order to ensure a successful analysis, the workflow comes with a few limitations though.

- All batches must be analyzed using the same spectral library and acquisition method.
- SNE files of directDIA experiments are currently not supported
- Regulation analysis and related processes (like Heatmap and Volcano plot) are currently not supported
- Final report exports are currently only supported in long-format (no Pivot report possibilities).
- Normalization is limited to the "Global Normalization" option
- Information on sample grouping will be merged from the individual SNE files' Condition Setups and cannot be changed in the SNE combine workflow

You can use the SNE combine workflow from the Pipeline perspective.

1. Click on "Set up a SNE combine Process" at the bottom left. The SNE combine wizard dialog will open. Select multiple SNE files (or a folder with SNE files) you want to combine.
2. The next step in the wizard allows you to change settings for the DIA analysis. You can define Quantification and Protein Inference settings as well as your output files (in the "Pipeline Mode" settings; Figure 41).

- Note: Not all analysis settings are available because the identification results from the individual SNE files are used, and the associated settings cannot be changed during SNE file combination. Make sure you select the correct settings when you initially generate the SNE files.

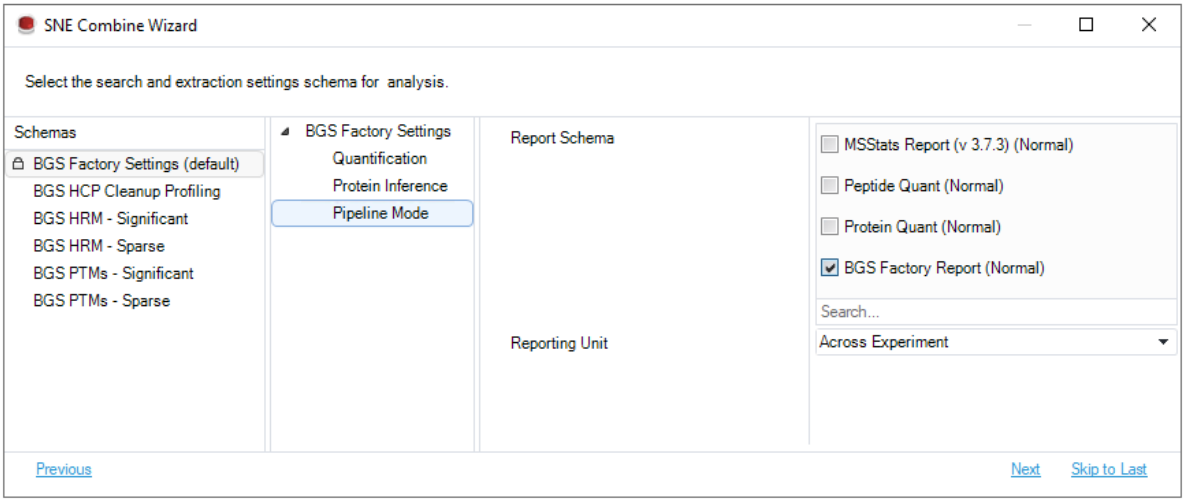


Figure 41. The SNE Combine Wizard guides through the parameters that are recalculated during multiple SNE files merging.

3. In the next step of the wizard, you can optionally select a protein database (FASTA file) that you would like to use for protein inference (instead of what was used for initial generation of the SNE files). This can be useful if the SNE files were made with different protein databases.
4. Finally, you need to define an output folder. The report files you selected will be saved there in a new subfolder (named after the experiment name with a time stamp). Once you click “Finish”, the SNE combine job will be added to your pipeline queue.

You can start the job by clicking “Run Pipeline”. Alternatively, you can first add additional SNE combine or DIA analysis jobs to the pipeline.

DIA Analysis Pipeline Mode Settings

Here you can choose which reports should be written, whether run-based or experiment-based reports should be performed, whether scoring histograms should be reported, and if the whole experiment should be saved to an *.sne file (Figure 42).

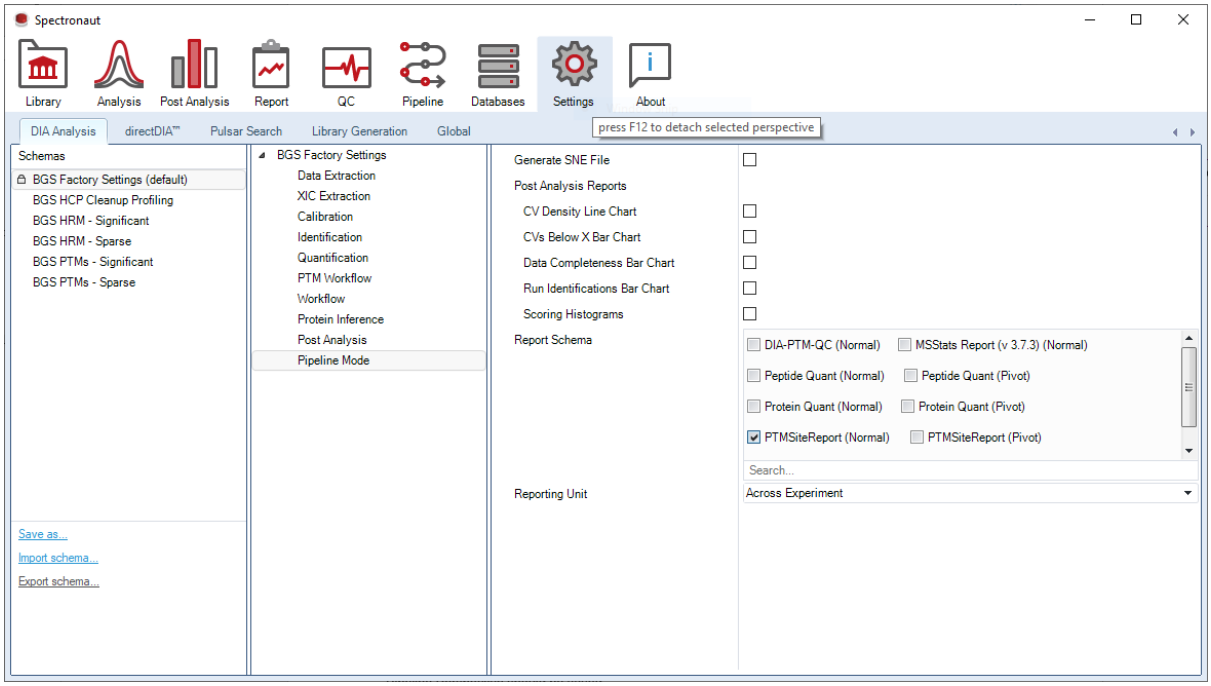


Figure 42. DIA Analysis Pipeline Mode Settings. Define which results from the Pipeline Perspective should be saved.

3.9 Databases Perspective

The Databases Perspective allows you to store and manage information that you will need to use when setting up analyses. This includes protein databases, gene annotations, peptide modifications, etc.

Protein Databases

This section of the settings perspective lets you import and manage your protein databases. Spectronaut uses protein databases (FASTA files) to make searches for library generation with Pulsar and to do protein inference. The protein databases contain all of the sequences, as well as meta-information extracted from the FASTA protein headers, using the specified parsing rule. Spectronaut already contains the UniProt parsing rule, but you can add a new rule by clicking "New Rule" in the Protein Databases page or during an importation (Figure 43).

In order to import a new proteome database from FASTA click on "Import…" in the bottom left corner" (Figure 43). While importing a new protein database from FASTA, Spectronaut will try to find the appropriate parsing rule for this file format from the already specified rules. Should no matching parsing rule found, you will be asked to specify a new one. Once your new protein database is imported it will be available in the Databases tree during the analysis set-up.

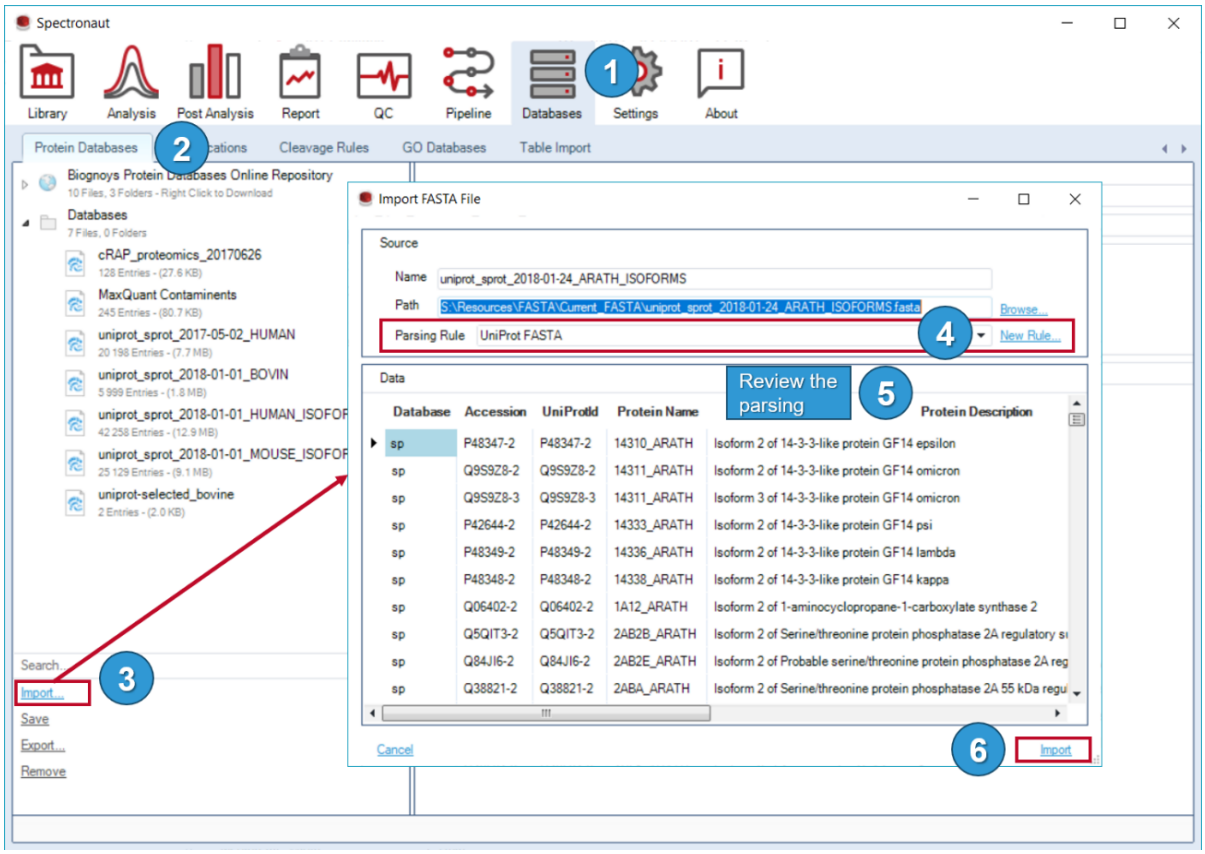


Figure 43. Importing a new FASTA file into Spectronaut. Your new database will appear in the Databases tree and will be available for setting-up analyses.

Modifications

The Spectronaut modifications panel provides a central location to manage modifications. Modifications contained here are used to automatically detect modified sequences in a search engine output when creating a spectral library from an external search engine. They are also used to detect modified sequences encoded in an external spectral library by means of synonyms. We recommend that modifications in a spectral library are mapped to Spectronaut modifications for optimal downstream analysis (see Section 3.3.1.3 and Figure 9). If a library is generated using Spectronaut, this is handled automatically.

Spectronaut comes with a database of default modifications for all search engines. If you use special modifications, please import the corresponding modifications file into Spectronaut.

3.9.1.1 Importing Modifications from Search Engine

To import non-default modifications into Spectronaut, you can batch import (see also Table 5):

- For MaxQuant, using the modifications.xml file in its bin folder.

- For ProteinPilot, using Unified Modification Catalog.xlsx, located in the ProteinPilot/Help folder in the Program Files.
- For Proteome Discoverer, no action is required.
- For Mascot, non-default modifications have to be created as custom modifications, see below.

When possible, Spectronaut will merge identical modification from multiple sources and save only the necessary search engine specific mapping information. However, if it is not able to unambiguously merge two or more modifications, you will be asked to resolve any conflicts at the time of import. You can tell if a modification has been mapped to multiple search engines by looking at the "Mapped to" data grid in the panel.

3.9.1.2 Creating custom modifications

It is possible to specify a new modification. This action has two main applications:

1. Incorporate modifications for Mascot searches which are not in the Unimod database (non-default ones).
2. Add a new label to generate a labeled library

To create a new modification, click "new" in the bottom left corner, give a name to your modification and click "OK" (Figure 44). Edit your new modification as desired and click "Save" in the bottom left corner. You can also modify an existing modification by clicking "Save As…".

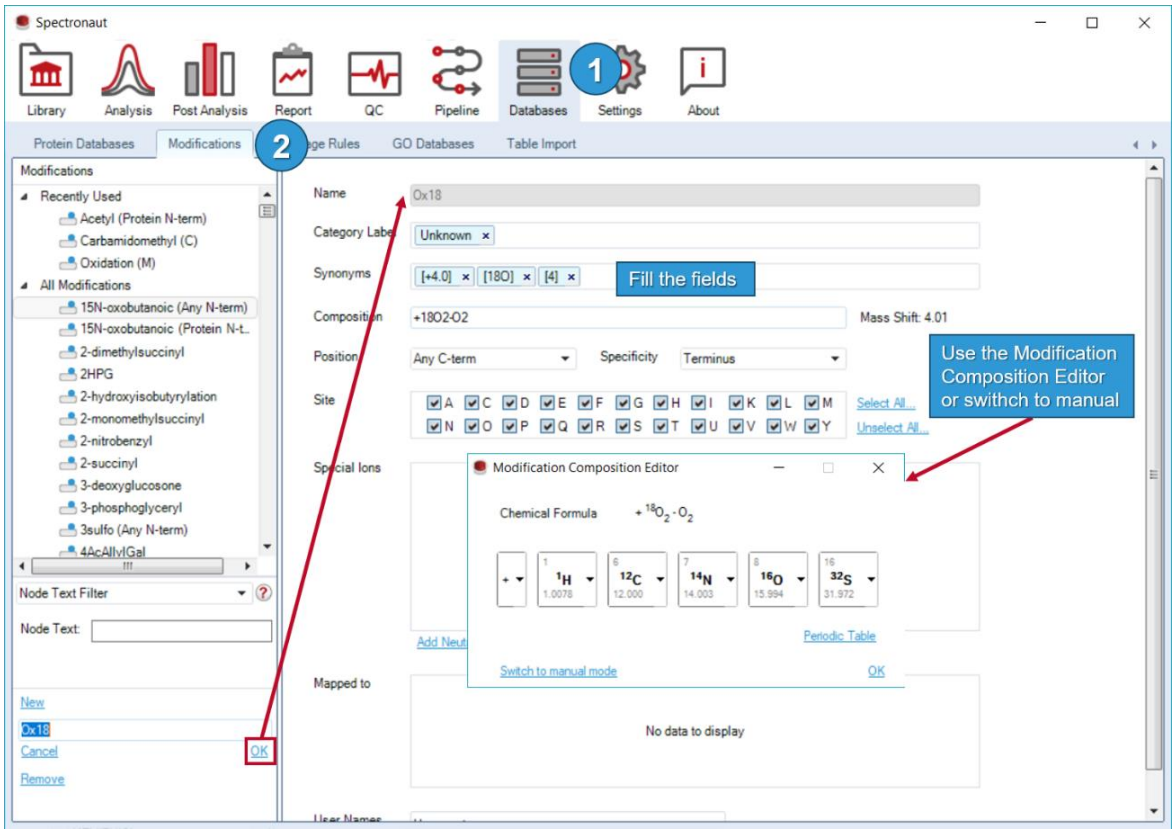


Figure 44. Adding a new modification to the database.

Cleavage Rules

This tool lets you define the rule to *in silico* digest your proteins from the protein database(s). Digest rules are applied whenever you do a Pulsar search (in library generation or in directDIA™). The most frequent rules are already included in Spectronaut, such as Trypsin, Trypsin/P, and LysC.

To design your own rule, you have to click on an existing one, modify it and click on "Save as…" in the bottom left corner (Figure 45). The rules are defined by which sites are cleaved by the enzyme. In the Digest Rule page, you will see a 20 x 20 matrix containing all possible combinations of amino acids. Select the combination where your enzyme cleaves (Figure 45). At the bottom you will see a preview of how a sequence will look like after being cleaved following your digest rule. You can also include a description.

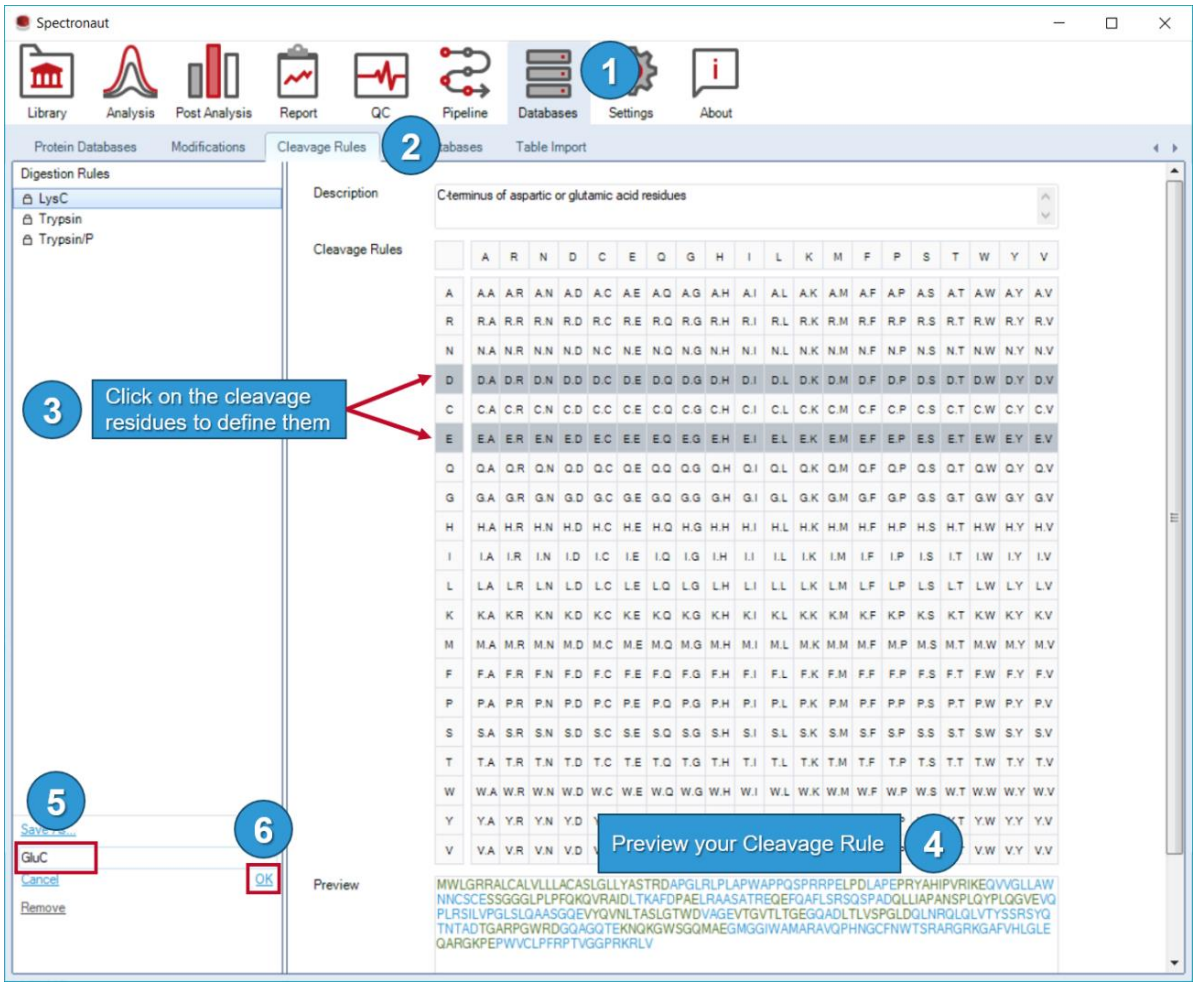


Figure 45. Define a new Cleavage Rule. The Cleavage Rule editor will allow the generation of new cleavages rules in a very friendly manner.

GO Databases

Similar to the Protein Databases, Gene Ontology (GO) Databases in Spectronaut are used to further annotate your data. This annotation will be used for calculating term enrichment and give further biological insight into the differential abundance results. The GO databases section manages two different data structures: Gene Ontologies and gene annotations. Find more details about each below.

3.9.1.3 Gene Ontologies

These allow you to import complex gene ontology structures in the shape of graphs. These structures are used for hierarchical grouping of functions, components, and processes. Currently, Spectronaut supports the *.obo file format from the GO Consortium. The go-basic.obo is already part of the Spectronaut installation. Information from a gene ontology tree can only be used in combination with an organism specific gene annotation file.

3.9.1.4 Gene Annotations

The gene annotation file functions as a link between the protein identifier (Uniprot accession number) and the GO tree. In its most basic form, the gene annotation file must feature two columns:

1. The protein id that specifies the UniProt accession number
2. GO-ID that specifies the unique identifier of each GO term. In case you don't have the GO-ID information, this column will have to be substituted by two: the unique name of the term and its corresponding namespace (function, component, or process).

Using this format, Spectronaut will connect the protein entries of your analysis via the GO-ID with the respective entries in the Gene Ontology to annotate your data further.

The official GO Consortium annotation file (*.gaf) is recommended, but you can specify a custom annotation file.

To import a new gene annotation file into Spectronaut, go to the GO Databases page of the Databases Perspective. Click "Import Gene Annotation…" and navigate to your *.gaf file. The GO annotation will automatically appear in your Gene Annotations tree.

Table Import

Spectronaut® can remember column names in user spectral libraries. Once you import a new library format into Spectronaut, it will ask you whether it should store novel synonyms for column header. You can manage those synonyms in the column recognition settings tab.

3.10 Settings Perspective

The Settings Perspective of Spectronaut® is meant to define custom settings schemas for any of the processes performed by the software. In this perspective, you will see a tab corresponding to each of these processes: DIA Analysis, Pulsar Search, directDIA™ and Library Generation (Figure 46). In addition, you can alter global settings of Spectronaut in the Global page (see below).

Detailed information regarding each setting option can be obtained by hovering the mouse over the label of a specific settings variable (Figure 46).

Make your own setting schema by modifying one of the predefined ones. Go throughout the nodes and edit the corresponding settings. Once you are done with the customization, click "Save as…" in the bottom left corner to give a name to your schema, and click "OK" (Figure 46). Your new schema will appear in the tree and it will be available to be chosen during the set-up of your next analysis. You can set your newly created schema as default Spectronaut settings by right clicking on the schema and choosing “set as default”. Finally, the setting schemas can be exported in different file formats (*.prop, *.txt). For DIA and directDIA, we have introduced JSON Settings file (*.json), for settings override and more flexible pipeline integration (see Section 3.10.1.6).

See the Appendixes for detailed information about the numerous settings within each process.

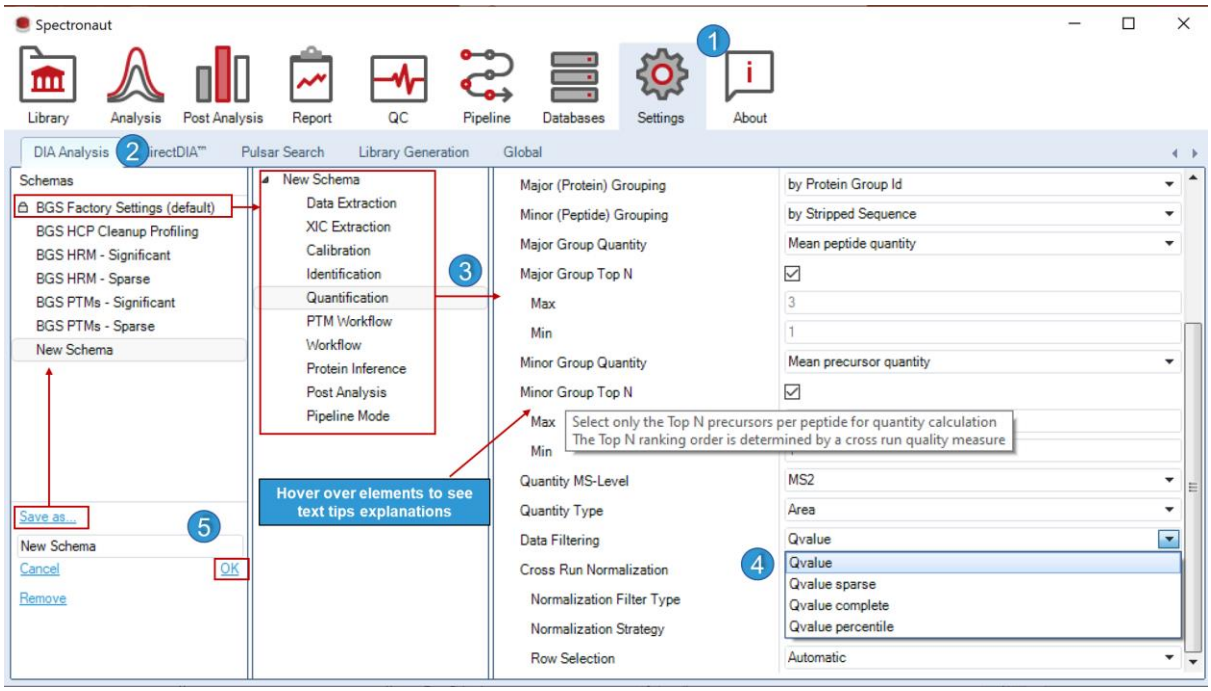


Figure 46. Make a custom schema for your analysis. The new schemas will be available during the subsequent analysis set-ups.

DIA Analysis Settings

The DIA Analysis Settings define the details of how Spectronaut should analyze the data, from DIA targeted data extraction to post-analysis calculations. These settings will specify important metrics, such as FDR cutoffs, decoy set generation to estimate scores, quantification settings, workflow to be used (label-free, labeled, spike-in), among many others. Find details of each setting in Appendix 1. DIA Analysis Settings (Section 7.1).

Pulsar Search Settings

These settings define how Pulsar should create the search-space when performing a search. You can specify the expected peptide characteristics (enzyme used, length, modifications, among others). Find details of each setting in the Appendix 7.2

directDIA™ Settings

The directDIA™ Settings are divided to section related to the Pulsar settings and section related to the quantitative DIA analysis. The Pulsar settings will define how to generate the search-space and how to perform the identification (FDR cutoffs). DIA analysis settings will define how the quantification will be performed. Find details of each setting in Appendix 3. directDIA™ Settings (Section 7.3).

Library Generation Settings

This set of settings defines the Library Generation process, either from Pulsar or from an external search engine. Metrics such as MS1 and MS2 tolerances, FDR cutoffs for identification confidence, peptide-based filters for your library, among others. Find details of each setting in Appendix 4. Library Generation Settings (Section 7.4).

Global Settings

The "Global" settings tab in the Settings perspective will allow you to change parameters that can be considered analysis unspecific. Here you will find options regarding plotting, working directories, as well as some general settings.

3.10.1.1 General

This section contains settings options that allow you to modify the default behavior of Spectronaut. For more information about these options use the tool-tip hover for each individual entry.

One important aspect of these settings is the File Name Parsing Strategy to let Spectronaut read information directly from the run file name. One of the most relevant uses of this function is the Condition Setup annotations. By defining the meaning of the different blocks in the file name, Spectronaut is able to obtain this information automatically (Figure 47).

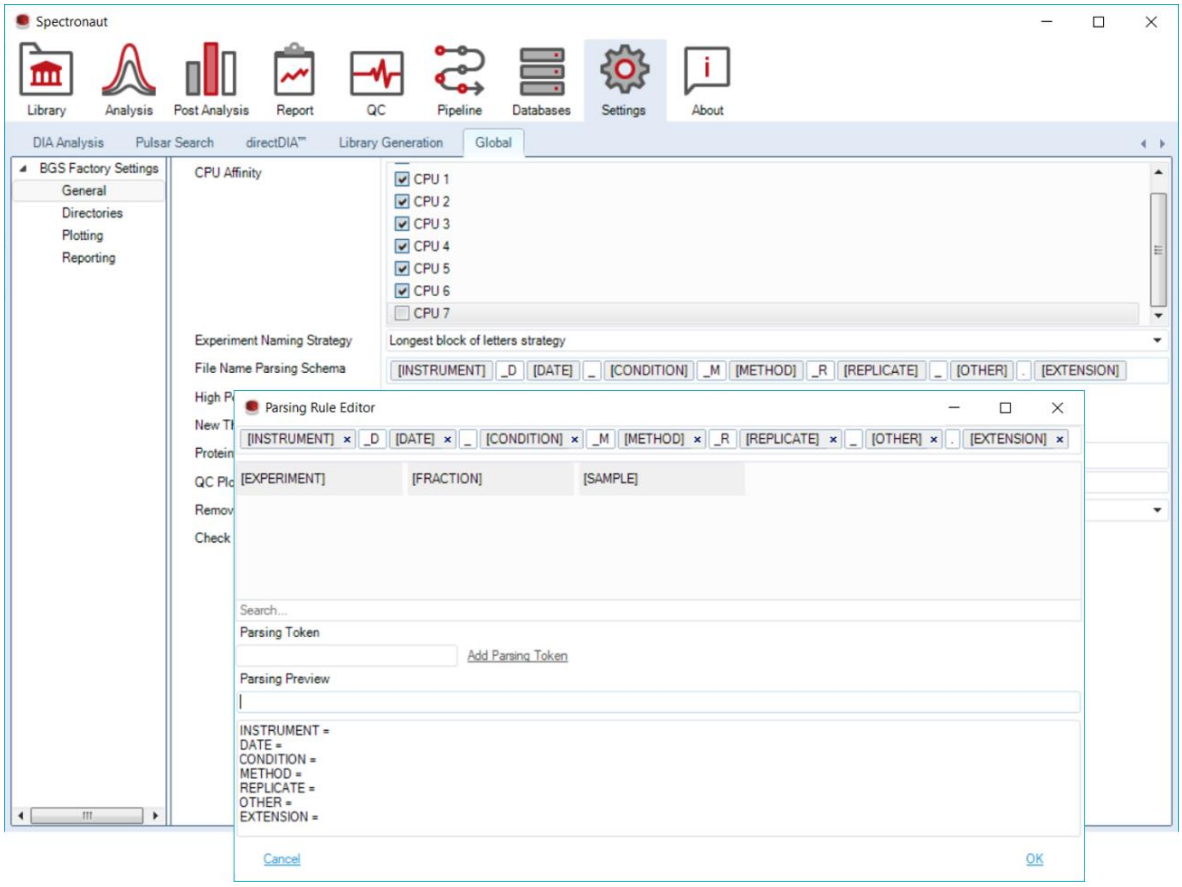


Figure 47. Parsing Rule Editor in the General Settings of the Global Settings page. Setting this rule properly will let Spectronaut read annotation information directly from the run file name.

3.10.1.2 Directories

Here you can setup the different storage paths for data managed by Spectronaut. Should you have a central storage location for all your DDA, you can specify this location here. This will allow Spectronaut to automatically map the correct shotgun acquisitions during the setup of the library generation pipeline (see Box 6). Please note that all changes within the "Directories" section will require a restart of Spectronaut in order to take effect. See some recommendation in Section 1.4.

3.10.1.3 Plotting

The plotting section allows you to customize the look and feel of most of the plotting options used in Spectronaut. You can specify whether XIC plots should show the integration

boundaries, as well as the expected elution time. Additionally, you can also apply smoothing to your plots. For more information about these options use the tool tip hover for each individual entry.

3.10.1.4 Reporting

Here you can specify two things:

1. Where to locate the results of the analysis performed via the Pipeline Perspective.
2. The default name of the files exported from Spectronaut.

Spectronaut Command Line Mode

Spectronaut can be run in Command Line Mode on both Windows (cmd or Powershell) and Linux operating systems. To get the overview of available commands and options please run "`spectronaut -h`" in where `spectronaut` is appropriate executable (`Spectronaut.exe` on Windows or `spectronaut` on Linux).

To activate Spectronaut license in the Command Line Mode, run "`spectronaut -activate <license_key>`". To deactivate the license, run "`spectronaut -deactivate`".

For Ubuntu Linux, use "`sudo dpkg -i spectronaut-VER.deb`" to install chosen version of Spectronaut. "VER" indicates version of installed Spectronaut. For the other Linux distributions, please follow the Read me file attached with a tarball installer.

Since version 18.5, options: `-lg`, `-direct`, `-convert`, `-combine` can be used also as commands without preceding dash sign (-). Moreover, every long option has its equivalent POSIX-compliant version preceded by double dash (--).

3.10.1.5 Library generation

When generating a library from the command line, call the `spectronaut` file followed by "`-lg`". The other options depend on the search engine of your choice.

3.10.1.5.1 Library Generation from Pulsar

When generating a library with the Pulsar search engine, start the command with

"`-lg -se Pulsar`". After that, use the following parameters to set up the experiment:

Table 8. Command line arguments for spectral library generation from Pulsar

Option	Explanation
`-se Pulsar`	Used as first command argument to run a library generation pipeline using Pulsar.
`-r` [path to file]	Adds a single raw file to the experiment. Any file format that is also supported during the analysis setup from the user interface is possible (*.raw, *.wiff, *.bgms, _HEADER.TXT, analysis.baf). This command can be used multiple times to add additional files.
`-d` [path to directory]	Adds all raw files, recognized by Spectronaut, from a given directory. This option includes vendor files that are already represented as folders (Bruker .d folders, Waters run folders). This command can be used multiple times to add additional directories.
`-sa` [path to file]	[OPTIONAL] Adds a specific Search Archive (*.psar file) to this library generation experiment. This command can be used multiple times to add additional files. The default location for Search Archives is configured in the Settings Perspective > Global > Directories.
`-sad` [path to directory]	[OPTIONAL] Adds all search archives (*.psar files) within a specified directory to this library generation experiment.
`-fasta` [path to file]	Specifies the path to a *.bgsfasta (Managed FASTA, namely FASTA file containing parsing rule i.e. a set of instructions that inform Spectronaut on how to read the column headers) file to be used as the DDA search space. This command can be used multiple times to add additional files. The default location for Managed FASTA files is configured in the Settings Perspective > Global > Directories.
`-rs` [path to file] OR `-rs` [schema-name]	[OPTIONAL] Specifies the Pulsar Search schema to be used for this search. If not specified, whatever is selected as the default schema will be used. This command can either be provided with a path to the schema file (*.prop) or with the schema name. The latter requires the schema to be in the internal Spectronaut Search schema repository (i.e., you should see it in the GUI).
`-es` [path to file] OR `-es` [schema-name]	[OPTIONAL] Specifies the Library Generation schema to be used for this search. If not specified, whatever is selected as the default schema will be used. This command can either be provided with a path to the schema file (*.prop) or with the schema name. The latter requires the schema to be in the internal Spectronaut Library Generation schema repository (i.e., you should see it in the GUI).
`-a` [path to file]	[OPTIONAL] Specifies the target location for the Search Archive file generated from this search. If not provided, the Search Archive will be stored in the default location. The default location for Search Archives is configured in the Settings Perspective > Global > Directories.

`-k` [path to file]	[OPTIONAL] Specifies the target location for the spectral library file generated from this search. If not provided, the library will be stored in the default location. The default location for Spectral Libraries is configured in the Settings Perspective > Global > Directories.
`-regex` [regular expression]	[OPTIONAL] Applies a regular expression as filter to the -d command.
`-n` [any text]	[OPTIONAL] Specifies the experiment name for this search. This name will be used to label the resulting spectral library and Search Archive. If not provided, Spectronaut will automatically generate an experiment name from the selected run file names.
`-inf` [path to file]	[OPTIONAL] Specifies the path to a *.bgsfasta (Managed FASTA) file to be used in this library for protein inference. The default location for Managed FASTA files is configured in the Settings Perspective > Global > Directories. By default, the FASTA files specified for the search space will be used for protein inference. This command can be used multiple times to add additional files.
`-go` [path to file]	Path to Gene Annotation file (could be multiple). This command can be used multiple times to add additional files.

3.10.1.5.2 Library Generation from External Search Engines

Table 9. Command line arguments for spectral library generation from an external search engine

Option	Explanation
`-lg -se <SearchEngineName>`	Used as first command argument to run a library generation pipeline using an external search engine. Replace <SearchEngineName> with one of the following choices: ProteomeDiscoverer, MaxQuant, ProteinPilot, BGSGenericSearchFormat, or Mascot.
`-sr` [path to search results]	Path to the search results. Please refer to the search engine specific section to see what type of results are needed for a given search engine.
`-rd` [path to raw files]	Specify the path to the run files.
`-o` [output library file]	Library file destination including the file name.
`-s` [path to file]	[OPTIONAL] Specifies the Library Generation schema to be used for this search result. If not specified, whatever is selected as the default schema will be used. This command should be provided with a path to the schema file (*.prop).

`-fasta` [path to file]	[OPTIONAL] Specifies the path to a *.bgsfasta (Managed FASTA) file to be used in this library for protein inference. The default location for Managed FASTA files is configured in the Settings Perspective > Global > Directories.

3.10.1.6 Spectronaut Analysis

In addition to the visual pipeline mode, Spectronaut is also capable of running the pipeline from command line. To run Spectronaut in command line mode you simply call the Spectronaut.exe file using the following parameter. From Spectronaut 18.5, you can run Spectronaut specifying either in search with a precompiled search library or search archive mode using "`spectronaut diaanalysis`" command, or library-free directDIA search mode by using "`spectronaut directdia`" command followed by options from table below.

Option	Explanation
`-h`	Displays help window with description of basics commands for running Spectronaut experiment.
`-r` [path to file]	Adds a single run file to the experiment. Any file format that is also supported during the analysis setup from the user interface is possible (*.htrms, *.raw, *.wiff, *.bgms, _HEADER.TXT, analysis.baf). This command can be used multiple times to add additional files.
`-d` [path to directory]	Adds all run files, recognized by Spectronaut, from a given directory. This option includes vendor files that are already represented as folders (Bruker: .d folders, Waters: .raw folders).
`-regex` [regular expression]	[OPTIONAL] Applies a regular expression as filter to the -d command.
`-a` [path to file]	Assigns a spectral library to every run in the experiment. This command can be used multiple times to add additional files.
`-ar` [path to file]	Assigns a spectral library to the last run added with the -r command. This command can be used multiple times to add additional files.
`-s` [path to file] OR `-s` [schema-name]	[OPTIONAL] Specifies the settings schema to be used for the analysis. If not specified, whatever is selected as the default schema will be used. This command can either be provided with a path to the schema file (*.prop) or with the schema name. The latter requires the schema to be in the default location for Spectronaut analysis schemas.

-setTemp <path>	[OPTIONAL] Change the temporary directory path.
-rs [path to schema]	[OPTIONAL] Specifies report schema.
-o [path to directory]	[OPTIONAL] Specifies the output directory for this experiment. All generated reports will be located in a sub-folder titled with the analysis date and the experiment name. By default, the results will be placed in the users specified default output location (global settings in Spectronaut) or in %Appdata%/Spectronaut/Results if nothing is specified.
-n [any text]	[OPTIONAL] Specifies the name of this experiment. If not provided, Spectronaut will automatically generate an experiment name from the selected run file names.
-fasta [path to file]	[OPTIONAL] Specifies the path to a *.bgsfasta file to be used in this experiment. In a library based (peptide-centric) analysis, this file is used for protein inference. This command is not optional in combination with the -direct command. This option can be used multiple times to add additional files.
-go [path to file]	[OPTIONAL] Specifies the path to the Gene Ontology (GO) annotation file to append. This command can be used multiple times to add additional files.
-direct	[OPTIONAL] Triggers a directDIA (spectrum-centric) pipeline instead of a library-based (peptide-centric) pipeline. Most commands work the same in this mode. However, the -fasta command is required while the -a and -ar commands will be ignored. This command has to be the first one to be called.
-con [path to file]	[OPTIONAL] Specifies the condition setup to be used for the post processing (such as regulation analysis). This condition setup file is best generated from the Spectronaut user interface and then exported for use in the command line mode. If not used, Spectronaut will use the specified run name parsing strategy to derive condition names for the samples.
-command [path to file]	[OPTIONAL] Specifies the path to a command (or arguments) file to be used instead of the standard command line arguments. This must be the first command if used. Any subsequent commands will be ignored. The command arguments file works the same as the regular command line arguments but is a line-based file that is read in instead. One can generate an example for the command arguments file from the last page of the Spectronaut experiment setup.
-rExt [path to file]	Adds a single run file for the 1-step hybrid library-based DIA-analysis. Any file format that is also supported during the analysis setup from the user interface is possible (*.htrms, *.raw, *.wiff, *.bgms, _HEADER.TXT, analysis.baf). This command can be used multiple times to add additional files.

-dExt [path to directory]	Adds all run files, recognized by Spectronaut, from a given directory for the 1-step hybrid library-based DIA-analysis
-sne <path>	Load experiments from .sne file.
-j [path to file]	Specifies a ".json" file for settings override and flexible pipeline integration

An example:

```
Spectronaut.exe –d "C:\data\My Experiment" –a "C:\data\My
Experiment\library.txt" –s "my_settings" –o "C:\data\My
Experiment\Results" –n "My Experiment 1" –f ".*\.wiff"
```

If you encounter problems with the automatic parsing of your spectral library, please first try to load the spectral library using Spectronaut's graphical user interface and make sure that all necessary columns are recognized automatically.

3.10.1.7 Combining SNE files

To execute SNE Combine pipeline in Command Line mode, run “spectronaut -combine” followed by options from table below:

Option	Explanation
-s <path_to_settings_schema>	Select the path to the settings schema file you want to use.
-n <name>	Specify the experiment name.
-fasta <fasta_file_path>	Specifies the path to a *.fasta or *.bgsfasta files to be used in this experiment for protein inference.
-sne <path_to_sne_file>	Specify *.sne files that you want to combine. This option can by used multiple times
-d <directory_path>	Specify all *.sne files that you want to combine by their directory. This option can by used multiple times.
-o <output_path>	Specify the directory where result will be put.

One can also export a command line setup from the last page of the experiment setup wizard for a DIA analysis in Spectronaut.

To get more information on command line mode, watch our short video tutorial here.

4 HTRMS Converter

Together with Spectronaut® installation, a complementary software called HTRMS Converter will be also installed.

The HTRMS Converter converts DIA run files into a Biognosys compatible format called HTRMS. These files are pre-processed and optimized to be analyzed in Spectronaut. Converting run files into HTRMS files is very useful if you need to analyze the same files several times. The overall analysis time will be significantly reduced.

The HTRMS converter is free to use and can be run on multiple computers without the requirement of a license key. To get more information on the use of the HTRMS converter, watch our short video tutorial.

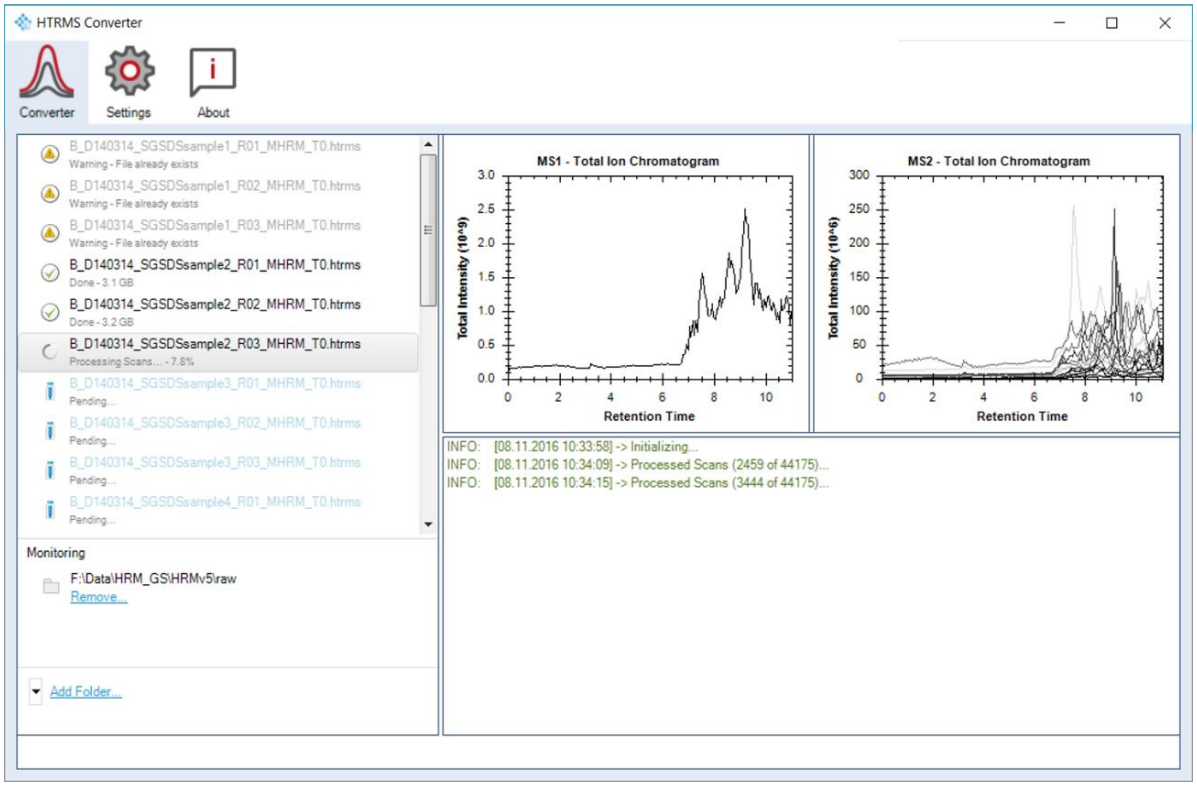


Figure 48. The new HTRMS converter with multiple tasks added to the task-list and monitoring a local directory for new MS/MS files.

4.1 File Conversion

In order to select one or more files from your hard drive to be converted, click on the drop-down arrow in the Converter perspective and select "Add Files…". After selecting one or more MS/MS files from the hard-drive, an input form will appear which allows you to specify the

conversion parameter. Click on "OK" to add all selected files to the main task list in the Converter perspective. You can add more tasks to the list at any moment.

The HTRMS Converter also allows the deconvolution of spectra, collected using overlapping windows, as proposed by (Amodei et al., 2019). This feature can be applied for varied- or fixed-size MS/MS scans. To enable this function, head to the HTRMS file settings and select the MS2 DeMultiplexing feature.

4.2 Folder Conversion

Using the drop-down arrow and selecting "Add Folder…" will ask you to specify an input folder. The folder conversion will automatically convert all valid MS/MS files within the target folder that meet some basic filter criterion.

Using the folder conversion, you now have access to the "Batch Conversion" settings, which allows you to specify filter criteria such as vendor or file age. You can also specify monitoring of the folder in order to automatically convert every new file that is added to the input folder.

4.3 HTRMS Converter Command Line Mode

Invocation on Windows:

```
HTRMSConverter.exe –i [PATH] –o [PATH] –s [mySettings] -nogui
```

The HTRMS conversion can be also invoked on Linux and Windows from Spectronaut by running "`spectronaut -convert`" followed by options from table below.

Table 10 Command line arguments for HTRMS Converter

Argument	Required	Explanation
`-i [path]`	TRUE	Path to a raw file, or a folder containing raw files
`-o [path]`	FALSE	Destination path for a HTRMS file (including file name) or path to destination folder. Default: raw-path but replace file ending with ".htrms"
`-s [schema]`	FALSE	Path or name of conversion settings schema to use Default: BGS_FactorySettings
`-nogui`	FALSE	Runs the conversion in command line window instead of starting the task in UI mode

5 BGMS Raw API

Besides the specified vendor formats, Spectronaut® also supports the Biognosys generic MS file format called BGMS. This file format can be generated using the BGSRawAPI.dll that is installed together with Spectronaut. This API is written in C# .NET and can be used with any Microsoft .NET language to build a custom file processor for raw vendor formats. This is especially useful if you want to process DIA data from an instrument that Spectronaut does not natively support or if your DIA method requires some special scan pre-processing not implemented in Spectronaut. An example project for how to use the BGMS raw API can be requested after reaching out us via Help Center.

6 References

Amodei, D., Egertson, J., MacLean, B. X., Johnson, R., Merrihew, G. E., Keller, A., Marsh, D., Vitek, O., Mallick, P., & MacCoss, M. J. (2019). Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *Journal of the American Society for Mass Spectrometry*, *30*(4), 669–684. https://doi.org/10.1007/s13361-018-2122-8

Bekker-Jensen, D. B., Bernhardt, O. M., Hogrebe, A., Martinez-Val, A., Verbeke, L., Gandhi, T., Kelstrup, C. D., Reiter, L., & Olsen, J. V. (2020). Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nature Communications*, *11*(1), 787. https://doi.org/10.1038/s41467-020-14609-1

Bekker-Jensen, D. B., Martinez-Val, A., Steigerwald, S., Rüther, P. L., Fort, K. L., Arrey, T. N., Harder, A., Makarov, A. A., & Olsen, J. V. (2020). A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Molecular & Cellular Proteomics*, mcp.TIR119.001906. https://doi.org/10.1074/mcp.TIR119.001906

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bilbao, A., Zhang, Y., Varesio, E., Luban, J., Strambio-De-Castillia, C., Lisacek, F., & Hopfgartner, G. (2015). Ranking Fragment Ions Based on Outlier Detection for Improved Label-Free Quantification in Data-Independent Acquisition LC-MS/MS. *Journal of Proteome Research*, *14*(11), 4581–4593. https://doi.org/10.1021/acs.jproteome.5b00394

Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinović, S. M., Cheng, L.-Y., Messner, S., Ehrenberger, T., Zanotelli, V., Butscheid, Y., Escher, C., Vitek, O., Rinner, O., & Reiter, L. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Molecular & Cellular Proteomics : MCP*, *14*(5), 1400–1410. https://doi.org/10.1074/mcp.M114.044305

Bruderer, R., Bernhardt, O. M., Gandhi, T., Xuan, Y., Sondermann, J., Schmidt, M., Gomez-Varela, D., & Reiter, L. (2017). Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Molecular & Cellular Proteomics : MCP*, *16*(12), 2296–2309. https://doi.org/10.1074/mcp.RA117.000314

Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W.-J., Webb-Robertson, B.-

J. M., Smith, R. D., & Lipton, M. S. (2006). Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *Journal of Proteome Research*, *5*(2), 277–286. https://doi.org/10.1021/pr050300l

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V, & Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *Journal of Proteome Research*, *10*(4), 1794–1805. https://doi.org/10.1021/pr101065j

Distler, U., Kuharev, J., Navarro, P., Levin, Y., Schild, H., & Tenzer, S. (2014). Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nature Methods*, *11*(2), 167–170. https://doi.org/10.1038/nmeth.2767

Dunn, O. J. (1961). Multiple Comparisons among Means. *Journal of the American Statistical Association*, *56*(293), 52–64. https://doi.org/10.1080/01621459.1961.10482090

Egertson, J. D., Kuehn, A., Merrihew, G. E., Bateman, N. W., MacLean, B. X., Ting, Y. S., Canterbury, J. D., Marsh, D. M., Kellmann, M., Zabrouskov, V., Wu, C. C., & MacCoss, M. J. (2013). Multiplexed MS/MS for improved data-independent acquisition. *Nature Methods*, *10*(8), 744–746. https://doi.org/10.1038/nmeth.2528

Escher, C., Reiter, L., MacLean, B., Ossola, R., Herzog, F., Chilton, J., MacCoss, M. J., & Rinner, O. (2012). Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics*, *12*(8), 1111–1121. https://doi.org/10.1002/pmic.201100463

Geiger, T., Cox, J., & Mann, M. (2010). Proteomics on an Orbitrap benchtop mass spectrometer using all-ion fragmentation. *Molecular & Cellular Proteomics : MCP*, *9*(10), 2252–2261. https://doi.org/10.1074/mcp.M110.001537

Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., Bonner, R., & Aebersold, R. (2012). Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Molecular & Cellular Proteomics*, *11*(6), O111.016717-O111.016717. https://doi.org/10.1074/mcp.O111.016717

Huang, T., Bruderer, R., Muntel, J., Xuan, Y., Vitek, O., & Reiter, L. (2019). Combining Precursor and Fragment Information for Improved Detection of Differential Abundance in Data Independent Acquisition. *Molecular &Amp; Cellular Proteomics*, mcp.RA119.001705. https://doi.org/10.1074/mcp.RA119.001705

Kiyonami, R. (2014). *Large Scale Targeted Protein Quantification Using WiSIM-DIA on an Orbitrap Fusion Tribrid Mass Spectrometer*.

Lambert, J.-P., Ivosev, G., Couzens, A. L., Larsen, B., Taipale, M., Lin, Z.-Y., Zhong, Q., Lindquist, S., Vidal, M., Aebersold, R., Pawson, T., Bonner, R., Tate, S., & Gingras, A.-C. (2013). Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nature Methods*, *10*(12), 1239–1245.

https://doi.org/10.1038/nmeth.2702

Li, W., Chi, H., Salovska, B., Wu, C., Sun, L., Rosenberger, G., & Liu, Y. (2019). Assessing the Relationship Between Mass Window Width and Retention Time Scheduling on Protein Coverage for Data-Independent Acquisition. *Journal of the American Society for Mass Spectrometry*, *30*(8), 1396–1405. https://doi.org/10.1021/jasms.8b06074

Meier, F., Geyer, P. E., Virreira Winter, S., Cox, J., & Mann, M. (2018). BoxCar acquisition method enables single-shot proteomics at a depth of 10,000 proteins in 100 minutes. *Nature Methods*, *15*(6), 440–448. https://doi.org/10.1038/s41592-018-0003-5

Mi, H., Muruganujan, A., Casagrande, J. T., & Thomas, P. D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature Protocols*, *8*(8), 1551–1566. https://doi.org/10.1038/nprot.2013.092

Moseley, M. A., Hughes, C. J., Juvvadi, P. R., Soderblom, E. J., Lennon, S., Perkins, S. R., Thompson, J. W., Steinbach, W. J., Geromanos, S. J., Wildgoose, J., Langridge, J. I., Richardson, K., & Vissers, J. P. C. (2018). Scanning Quadrupole Data-Independent Acquisition, Part A: Qualitative and Quantitative Characterization. *Journal of Proteome Research*, *17*(2), 770–779. https://doi.org/10.1021/acs.jproteome.7b00464

Silva, J. C., Gorenstein, M. V, Li, G.-Z., Vissers, J. P. C., & Geromanos, S. J. (2006). Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Molecular & Cellular Proteomics : MCP*, *5*(1), 144–156. https://doi.org/10.1074/mcp.M500230-MCP200

Supek, F., Bošnjak, M., Škunca, N., & Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS One*, *6*(7), e21800. https://doi.org/10.1371/journal.pone.0021800

Tsou, C.-C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A.-C., & Nesvizhskii, A. I. (2015). DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. *Nature Methods*, *12*(3), 258–264, 7 p following 264. https://doi.org/10.1038/nmeth.3255

Zhang, B., Chambers, M. C., & Tabb, D. L. (2007). Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *Journal of Proteome Research*, *6*(9), 3549–3557. https://doi.org/10.1021/pr070230d

7 Appendixes

7.1 Appendix 1. DIA Analysis Settings

XIC Extraction

Basic Settings

The XIC IM Extraction Window defines if ion mobility should be used to predict the elution of a peptide:

- Dynamic (default setting): Spectronaut will dynamically adjust the XIC window in Ion Mobility dependent manner, based on large sample set during calibration. A correction factor can be applied, e.g., a factor of 2.0 would mean that you want to use 2 times the window that Spectronaut suggests. The default settings are recommended for most applications.
- Full: Spectronaut will use the full ion mobility space to find the target.

The XIC RT Extraction Window defines if iRT should be used to predict the elution of a peptide:

- Dynamic (default setting): Spectronaut will determine the ideal extraction window dynamically depending on iRT calibration and gradient stability. Sections of the gradient that show higher variability during the calibration step will automatically be extracted using wider windows.
- Static: Spectronaut will use a fixed width (in min).
- Full: Spectronaut will use the full gradient width to find the target.

Advanced Settings

MS Mass Tolerance Strategy: Spectronaut® will calculate the ideal mass tolerances for data extraction and scoring based on its extensive mass calibration. However, you can also specify your preferred tolerances for both MS1 and MS2 levels. You can choose amongst:

- Dynamic (default): determined by Spectronaut based on the extensive mass calibration. You can set a correction factor (default is no correction = 1)
- Relative: set a relative mass tolerance in ppm of the target ion mass
- Static: set a fix mass tolerance in Thompsons.

Calibration

Basic Settings	
Calibration Mode (**only relevant when using HTRMS files**)	Determines whether a re-calibration of the selected HTRMS files should be triggered or not • **None**: keep current stored calibration. • **Automatic** (default): Spectronaut will decide whether a recalibration is needed or not. • **Force**: will recalibrate the HTRMS files.
MZ Extraction Strategy	Defines which MS peak within the m/z tolerance will be picked during mass calibration • **Maximum Intensity** (default): use the data point with highest intensity value within the calibration training tolerance. • **Nearest Peak**: Pick the peak closest to the theoretical m/z within the calibration training tolerance. This is useful in labeled experiments with extremely narrow mass differences (<50 ppm) between the channels.
Used Biognosys' iRT Kit (recommended)	If the iRT Kit is in the sample, select this option for Spectronaut to expect the iRT Kit in the DIA data during calibration. Recommended to be selected by default in most cases, and only deselected if experiencing issues with the iRT calibration.
Advanced Settings	
Allow source-specific iRT Calibration	Enable or disable source-specific iRT Calibration. If selected (default) Spectronaut will generate libraries containing as many iRT values as different sources exist in the dataset. (Box 7).

Precision iRT	If selected and available, Spectronaut will use a larger set of calibration peptides to perform an extensive calibration and improve iRT precision and accuracy. Exclude Deamidated Peptides (default and recommended): will exclude all peptides from the spectral library that contain deamidation from the calibration process. iRT ←→ RT Regression Type: • **Linear Regression**: will do a linear iRT calibration **Local (Non-linear) Regression**: this allows to calibrate non-linear gradients, but also to correct for possible variations of a linear gradient. If Spectronaut does not find enough data points, it will automatically change to linear calibration.
Calibration Carry-Over	Recommended only for Host-Cell Proteomics (HCP) workflows. Spectronaut will take the calibration from the best performing run in the experiment and will apply this calibration to the rest of the runs. Consider using a higher XIC extraction width factor.

MS1 Mass Tolerance Strategy	Allows to determine MS1 tolerance for XIC extraction and scoring during calibration. One could choose from the following three options: • System Default – Spectronaut chooses MS1 tolerances specific for the different mass analyzers, namely: • Any Orbitrap +/- 40ppm • Any TOF (other than Bruker and Waters) +/- 40ppm • Bruker TOF +/- 40ppm • Waters TOF +/- 80ppm • Any Ion Trap +/- 0.5th • Relative – target mass tolerance is selected in relative units – ppm Static – a fixed mass tolerance is selected in Thomsons.
MS2 Mass Tolerance Strategy	• Allows to determine MS2 tolerance for XIC extraction and scoring during calibration. The option one can choose are the same as described above for MS1 Mass Tolerance Strategy.

Identification

Basic Settings	
Precursor PEP Cutoff and Protein PEP Cutoff	Specify the Posterior Error Probability (Precursor PEP and Protein Group PEP) cutoff that should be considered identified. This primarily affects which precursors and proteins are quantifiable. When using either the global or run-wise imputation strategy, precursors and protein groups that do not satisfy their respective PEP cutoff will be imputed.
Precursor Qvalue Cutoff and Protein Qvalue Cutoff	Choose your q-value (FDR) cutoff on precursor and protein level. Only those passing the cutoff will be considered identified and use for other subsequent processes. The protein Qvalue cutoff should be specified on both: experiment and run-wise level.
Single Hit Definition	Define what should be considered a protein single hit: stripped sequence, modified sequence, or peptide precursor ID. If only a single instance of the selected definition is identified across the entire experiment, the protein group will be marked as a single-hit protein group.
Exclude Single Hit Proteins	Discard protein groups identified with only one peptide hit (as defined above).
Advanced Settings	
Exclude Duplicate Assays	Spectronaut will keep only the best performing assay if a peptide is duplicated in the libraries.

Generate Decoys	• If unchecked, decoys have to be provided *a priori* in the library for Spectronaut to estimate the q-values (Qvalue). In such case the decoys need to be annotated in the column "**IsDecoy**" with the label "TRUE", and the targets need to be annotated with the label "FALSE". • If checked, you can specify the following: Decoy Method: defines how to generate the decoys. For details, please use the text hovers in the software. The default option is Mutated. Furthermore, since Spectronaut 14, the setting of Preferred Fragment Source was introduced. It could be chosen between two options: Template Fragments (will carry over the fragmentation pattern from the template peptide and just recalculate the masses based on the new sequence; recommended when other than HCD or CID fragmentation is used) and NN Predicted Fragments (neural network based strategy; will generate the ideal fragmentation pattern based on the newly generated decoy sequence). Decoy Limit Strategy: set the maximum number of decoys to be generated: • **Dynamic**: specify the number of decoys a fraction of the number of targets. • **Static**: choose a decoy limit as a fixed number of decoys. **None**: generate the same number of decoys as targets.
Machine Learning	• Per Run (default): calculates the discriminant scores (Cscores) and q-values (Qvalues) per run. • Across Experiment: makes a experiment-wise Cscore space. Can compromise the sensitivity.
Pvalue Estimator	Specify how you prefer the null distribution to be estimated to calculate the p-values: Kernel Density or Normal Distribution Estimator.

Quantification

Basic Settings	
Precursor Filtering	Decide how to apply the q-value filter on the precursors in an experiment-wide manner to quantify protein groups: • **Identified (Qvalue)** (default): only those precursors passing the q-value cut-offs will be reported (considered as quantified) and, accordingly, used for statistical testing of differential abundance. Those cases not passing the cut-off will be tagged as "Filtered". By the default there will be no imputation of missing values. • **Identified in All Runs** (complete): the peptide precursor needs to pass the q-value threshold in all the samples to be reported. This is the most stringent filter and produces the smallest data matrix. • **Identified in % of Runs:** the peptide precursor needs to pass the Qvalue threshold in the user defined percentage of all experimental samples. For instance, if you set a 50th percentile cutoff (0.5), the peptide precursor needs to pass the Qvalue in 50% or more of your samples to be reported. In the samples where it was not identified below the significance threshold (default ≤ 0.01), the specified Imputing Strategy will be applied. Imputing strategy: the imputing strategy defines how to estimate the missing values (identifications not fulfilling the FDR threshold). • **Use Empirical Noise:** the best real picked signal will be reported. • **Global Imputing**: missing values are imputed based on a random sampling from a distribution of low abundant signals taken across the entire experiment. • **Run Wise Imputing**: missing values are imputed based on a random sampling from a distribution of low abundant signals taken within the corresponding run. This is useful for large scale experiments.
Proteotypicity Filter (only with automatic inference)	Choose whether you want to quantify only based on non-shared peptides, either at the level of protein (very stringent) or at the level of protein group.

Protein LFQ method	Specify how protein level label-free quantification should be performed. • **Automatic** will pick MaxLFQ for smaller experiments (<= 500 Runs) and Quant2.0 for experiments exceeding this cutoff. Such setting is recommended due to exponential run-time behavior of MaxLFQ algorithm. • **MaxLFQ** derives label-free quantities based on inter-run peptide ratios. **Quant 2.0 (SN standard)** is based on aggregation of minor group quantities based on user specified TopN and summation strategies
Quantity MS-Level	Choose which MS level you want to use to perform quantification: MS1 or MS2. MS2 level = sum of the quantities of the number of fragment ions per precursor, as specified in the spectral library.
Quantity type	Decide which feature of the peaks should be used for quantification: area under the curve within integration boundaries or apex peak height.

Cross Run Normalization	Apply label-free normalization to the whole dataset. Choose among the available Normalization Strategies: • **Automatic**: peaks between local and global normalization based on the number of runs (for n<500, the local normalization will be performed). • **Global Normalization**: choose to use median, average or geometric mean. • **Local Normalization**: based on the Local Regression Normalization described by Callister *et al.* 2006. **Normalization Filter Type** allows to limit the selection of peptides that are used for the normalization by specifying a filter: • **None** will use no specific filters for selecting peptides to use for normalization • **Library Name** will select peptides based on the name of the library • **FASTA Name** will select peptides based on the name of the associated parent FASTA file • **Modification Type** will allow selection of peptides for the normalization based on their modification state. **Row Selection**: choose which precursor profiles to use for normalization: • Automatic: 10000 precursor profiles, ranked by level of completeness. • Qvalue sparse (see explanation in the Data Filtering options). • Qvalue complete (see explanation in the Data Filtering options). Qvaule percentile (see explanation in the Data Filtering options).
<u>Advanced Settings</u>	
Interference Correction	Exclude fragment ions detected as interferences across all runs (Bilbao et al., 2015). If checked, set a minimum number of features to be kept at MS1 and MS2 levels in order for Spectronaut to still perform the quantification. Exclude All Multi-Channel Interferences: for labeled experiments, exclude the fragment ions that are shared between heavy and light channels and would be extracted from the same scans.

Major (Protein) Grouping	Specify what should be considered as a protein (Protein Group Id or Gene Id).
Minor (Peptide) Grouping	Specify what should be considered as a peptide (Stripped Sequence, Modified Sequence or Precursor).
Major Group Quantity	Specify how you want the minor groups (peptides) to be used to calculate the major group (proteins) quantities.
Major Group Top N	Use the best N minor group elements to calculate the major group quantities. The elements are ranked by evidence count and quantity.
Minor Group Quantity	Specify how you want the precursors to be used to calculate the minor groups (peptides) quantities.
Minor Group Top N	Use the best N precursors to calculate the minor group quantities. The precursors are ranked by evidence count and quantity.

PTM Workflow

Here you can specify settings related to the PTM analysis

Basic Settings	
PTM localization	Calculates a PTM localization probability for all variable modification site options. A specified probability cut-off can be applied (default is 0.75).
PTM Analysis	Performs a PTM focused data analysis: • **Hierarchical Clustering** specifies whether or not to run hierarchical clustering also on PTM sites. • **Multiplicity** will specify whether to differentiate singly from multiply modified peptides during the site collapse. If checked, doubly and triply modified peptides will generate separate collapsed site objects to be analyzed during regulation analysis. This only applies to modifications of the same type.

Advanced Settings	
PTM Analysis	**Flanking Region** specifies how many amino acids should be considered for the “PTM.FlankinRegion” report field. **PTM consolidation** specifies how to derive quantity from set of parent peptides carrying a particular modification on a given modification site: • **Sum** would summarize quantities of all parent peptides that are carrying particular modification on a given modification site. • **Linear model** would firstly impute missing values for each parent peptide based on quantities reported in other runs. Afterwards, it would summarize quantities of all parent peptides as above.

Workflow

Here you can specify if you are running a label-free analysis or a different kind of quantification

Basic Settings	
Multi-Channel Workflow Definition	Specifies what kind of workflow should be used for multi-channel peptide assays. • **From Library Annotation**: will take the definition annotated in the library. Any precursors containing only one channel specification will automatically be treated as label-free. You can set what is the fallback option if there is no workflow annotation. • **Labeled**: peak detection and scoring will be applied to all channels. Quantification in post-analysis will be performed on the light to heavy ratio. • **Spike-in**: peak detection will be performed on only the reference (heavy) channel. Scoring and identification will be performed on the target (light) channel. The heavy channel is expected to be easily detectable and considered a peak-picking aid in this experiment. Quantification in post-analysis will be performed on the target to reference ratio. • **Force Label free**: will treat all multi-channel assays as label free. Peak detection, scoring and identification are applied in all channels. **Fallback Option**: If there is no workflow annotation for a given multi-channel peptide assay, it should be treated using this fallback option, either as Labeled or Spike-in.

Profiling Strategy	The profiling workflow allows the user to carry over the measured iRT of peptides that could be identified (q-value ≤ 0.01) in certain runs in order to fix integration boundaries in runs where an identification could not be achieved: **Template Correlation Profiling**: takes the best peptide signal in all runs as a template to find low abundant signals in the rest of the runs. **iRT Profiling**: takes the best peptide signal in all runs as a template and translates the empirical iRT to the integration boundaries of the low abundant signals in the rest of the runs. Profiling Row Selection: choose a row-wise q-value threshold for to choose which rows to profile (Minimum, Average or none) with a specified q-value threshold. Profiling Target Selection: specify which precursors should be readjusted (non-identified precursor or automatic).
<u>Advanced Settings</u>	
In-Silico Library Optimization	Changes a few parameters in the calibration and machine learning processes to optimize the DIA analysis towards very large (>800k precursors) spectral libraries with very low expected recovery rate. This is typically the case for In-Silico predicted spectral libraries made from a theoretical FASTA digest.
Unify Peptide Peaks Strategy	Unify the peak picking across different charge states of the same modified peptide based on the highest scoring instance.

Protein Inference

Spectronaut is able to perform protein inference using the IDPicker algorithm (Zhang et al., 2007). Protein grouping will be well defined and protein group counts will be comparable across search engines and spectral libraries. Spectronaut also checks which peptides are proteotypic. The options are:

- **Automatic**: (default) When using a Spectronaut formatted library (.KIT) and all information is available, Spectronaut will use the protein inference parameters (the FASTA file used and the specified protein cleavage rules) to re-calculate the protein inference based on all identified peptides in the experimental DIA evidence.
- **From Search Engine**: will keep the protein inference as defined by the search engine within the provided spectral library. The protein entries will not be re-grouped based on the experimental DIA evidence.
- **From protein-db matching**: you can overwrite the existing grouping by choosing a FASTA sequence database and specifying the protein cleavage rules.

Post-Analysis

Basic Settings	
Differential Abundance Testing	Choose whether you want Spectronaut to perform differential abundance testing (paired or unpaired Student's t-test) or not.
Differential Abundance Grouping	Select what the biological unit you want your results to be based on as defined in the quantification settings: Major (proteins) or Minor (peptides) Group. **Smallest Quantity Unit**: the unit on which to perform statistical testing. The default unit used for paired and unpaired t-test is protein group.**Use All MS-Level Quantities**: regulation analysis with combined MS1 + MS2 statistical model (Huang et al., 2019). The publication is to reference the combined MS1 + MS2 model for statistical testing, not the statistical test itself.
Calculate Explained TIC	The explained TIC is the proportion of the TIC that can be associated to identified peaks. Choose if and how the relative explained TIC is calculated. • **None**: Will not calculate the explained TIC chart • **Extensive**: Will use an extensive feature detection method to calculate the most accurately for the explained TIC plot **Quick**: will use a simple and fast feature detector and correct the results based on a heuristic technique.
Calculate Sample Correlation Matrix	Choose whether you want to calculate the sample correlation matrix. If selected, a new plot will be available in the Post Analysis perspective, Analysis overview node, called Sample Correlation. For large experiments it can be very time consuming.
Gene Ontology	Select which Gene Ontology database to be used. The default is the "GO Consortium go-basic" database. More options will appear once GO files (*.obo) are uploaded in the Database Perspective → GO databases.
Hierarchical Clustering	Choose whether you want to cluster your samples and potential candidates.
Advanced Settings	

Differential Abundance Testing	• **Assume Equal Variance** specifies if the unpaired two sample t-test should assume equal variance between the test groups • **Group-Wise testing Correction**: Perform the multiple testing correction on groups of IDs within compared conditions. If not selected, the multiple testing correction will be performed on experiment level over all conditions. • **Log2 Ratio Candidate Filter** defines default log2 ratio filter applied to the candidates-grid in the post-analysis perspective. • **Confidence Candidate Filter** defines the default confidence type and cutoff to apply to the candidates-grid in the post-analysis perspective.
Hierarchical Clustering	Defines further how Hierarchical Clustering should be performed.

Pipeline Mode

These settings are only relevant when running analyses from the Pipeline Perspective or from command line.

Specify if you want your analysis to be saved (with or without ion traces), which reports should be generated and saved, etc.

7.2 Appendix 2. Pulsar Search Settings

Configure the conditions for Pulsar to perform the search:

Peptides

Specify settings related to the peptide sequences.

Digest Type	**Specific**: both N- and C-terminus follow the specified digest rules **Semi specific (Free N-terminus)**: only the C-terminus follows the specified digest rules **Semi specific (Free C-terminus)**: only the N-terminus follows the specified digest rules **Semi-specific**: only one of the termini follows the specified digest rules **Unspecific**: no digest rules
Enzyme/Cleavage Rules	Proteases used to *in silico* digest the proteins from the protein database(s). Defined in Databases → Cleavage Rules
Maximum Peptide Length	Maximum number of amino acids allowed for a peptide
Minimum Peptide Length	Minimum number of amino acids allowed for a peptide
Missed Cleavages	How many consecutive cleavage sites the protease could miss
Toggle N-terminal M	Pre-processing of the protein database by toggling (processing both with and without) the protein N-terminal methionine (when there is one) to account for N-terminal methionine excision.

Labeling

Labeling applied	If selected, there are up to 3 channels where one can specify which labels, from the modifications database, are in each channel.

Applied modifications

Maximum Variable Modifications	Maximum number of variable modifications allowed to happen in one peptide at the same time. The higher this number, the more possible combinations, and the bigger the search space and the longer the calculation time
Select Modifications	**Fixed:** the amino acid always contains the modification **Variable**: the amino acid might or might not be modified. The more variable modifications, the bigger the search space and the longer the calculation time

7.3 Appendix 3. directDIA™ Settings

The directDIA™ Settings consist of settings related to Pulsar search and library generation and settings related to DIA Analysis..

directDIA Pulsar Search Settings

Peptides

Specify settings related to the peptide sequences.

Enzyme/Cleavage Rules	Proteases used to *in silico* digest the proteins from the protein database(s). Defined in Databases → Cleavage Rules
Digest Type	**Specific**: both N- and C-terminus follow the specified digest rules **Semi specific (Free N-terminus)**: only the C-terminus follows the specified digest rules **Semi specific (Free C-terminus)**: only the N-terminus follows the specified digest rules **Semi-specific**: only of the termini follows the specified digest rules **Unspecific**: no digest rules
Maximum Peptide Length	Maximum number of amino acids allowed for a peptide

Minimum Peptide Length	Minimum number of amino acids allowed for a peptide
Missed Cleavages	How many consecutive cleavage sites the protease could miss
Toggle N-terminal M	Pre-processing of the protein database by toggling (processing both with and without) the protein N-terminal methionine (when there is one) to account for N-terminal methionine excision.

Modifications

Maximum Variable Modifications	Maximum number of variable modifications allowed to happen in one peptide at the same time. The higher this number, the more possible combinations, and the bigger the search space and the longer the calculation time
Select Modifications	**Fixed**: the amino acid always contains the modification **Variable**: the amino acid might or might not be modified. The more variable modifications, the bigger the search space and the longer the calculation time

Identification

PSM FDR	Specify the FDR threshold on PSM level
Peptide FDR	Specify the FDR threshold on peptide level
Protein Group FDR	Specify the FDR threshold on protein level
PTM Localization Filter	Calculates a PTM localization probability for all variable modification site options. A specified probability cut-off can be applied (default is 0.75).

Tolerances

Spectronaut® will, by default, calculate the ideal mass tolerances to generate the library. Spectronaut performs two calibration searches: based on the first-pass calibration (rough calibration), the ideal tolerance for the second-pass calibration is defined; based on the second-pass calibration (finer calibration), the ideal tolerance for the main search is defined. Spectronaut will do this under default settings (Dynamic).

However, Spectronaut allows you to set you preferred tolerances for the different MS instruments (Thermo Ion Trap, Thermo Orbitrap, TOF). Hence, for both the calibration search (second-pass, finer calibration), and the main search, you can define your tolerances:

- **Dynamic**: determined by Spectronaut based on the precedent search (default). You can set a correction factor for MS1 and MS2 levels (default is no correction = 1)
- **Relative**: set a relative mass tolerance in ppm for both MS1 and MS2 levels
- **Static**: set a fix mass tolerance in Thomson for both MS1 and MS2 levels

Workflow

Fragment Ion Selection Strategy: defines the strategy to be used for selecting the top N fragment ions per peptide precursor

- **Intensity Based**: Prioritizes fragment ions by their intensity in the consensus spectra
- **Evidence Based**: Prioritizes fragment ions by how often they have been observed in the experiment for the same precursor
- **Maximize sequence coverage**: Groups fragment ions by type and position and ranks them based on the best fragment ion per group (either by intensity of by evidence) in an iterative manner.

In-Silico Generate Missing Channels: if generating a library for an isotopically labeled sample, you can activate this option to *in-silico* generate an assay for a missing label pair. For example, if you have a sample of only spiked-in heavy aqua peptides, it will make a light/heavy library where the label-free channel is generated *in-silico* based on the heavy channel and with a consistent fragment ion selection. This option works with n-channel search results as long as you are working with isotopic labels. The workflow options are:

Spike in workflow: will create a light channel for all heavy (SIS) peptides that are identified without a light counterpart.

Labeled workflow: will detect the labeling setup of an experiment and add the channels that are missing for a given peptide

Inverted spike-in: will create a heavy channel for all light peptides that are identified.

Use DNN Ion Mobility decides if ion mobility should be predicted based on deep neural network (DNN) for library generation.

- **Auto** will always predict Ion Mobility during library generation. Only if empirical ion mobility value is not available for the peptide, a predicted value will be used.
- **Always use predicted Ion Mobility** – a library will contain predicted ion mobility values, regardless if empirical information is available or not.
- **Never predict Ion Mobility** – ion mobility prediction will not be performed. Library will contain only empirical ion mobility values (if available).

Result Filters

You have a number of options to filter the search engine results. There are filters at the level of fragment ion and at the level of precursor. The filters are quite self-explanatory. Please, use the hover text-tools if you need more information. Find below some of the most relevant.

Fragment Ions	Filter peptides not fulfilling the conditions specified regarding fragment ions. Find more details by hovering over the option in the software. You can specify your defined criteria for: • Amino acid length • Ion charge • Ion type (y, b) • m/z window • Relative intensity (%)
Precursors	Amino Acids: Filter peptides containing specified amino acids Best N fragments per Peptide: specify the range in number of fragment ions based on response Best N Peptides per Protein Group: keep only the N most abundant peptides per protein Channel Count: keep peptides that are observed in a range of channels FASTA matched: keep only peptide that are found in a user-specified FASTA sequence database and abide the digest rules. Only available if protein inference is selected. Missed Cleavage: keep peptides with N or less missed cleavages Modifications: Filter peptides according to modifications. Find more details by hovering over the option in the software Peptide Charge: keep peptides with Proteotypicity: If a peptide has a unique protein mapping (matched to a single FASTA entry from the sequence database).

directDIA DIA Analysis settings

XIC Extraction

Basic Settings

The XIC IM Extraction Window defines if ion mobility should be used to predict the elution of a peptide:

- Dynamic (default setting): Spectronaut will dynamically adjust the XIC window in Ion Mobility dependent manner, based on large sample set during calibration. A correction factor can be applied, e.g., a factor of 2.0 would mean that you want to use 2 times the window that Spectronaut suggests. The default settings are recommended for most applications.
- Full: Spectronaut will use the full ion mobility space to find the target.

The XIC RT Extraction Window defines if iRT should be used to predict the elution of a peptide:

- Dynamic (default setting): Spectronaut will determine the ideal extraction window dynamically depending on iRT calibration and gradient stability. Sections of the gradient that show higher variability during the calibration step will automatically be extracted using wider windows.
- Static: Spectronaut will use a fixed width (in min).
- Full: Spectronaut will use the full gradient width to find the target.

Advanced Settings

MS Mass Tolerance Strategy: Spectronaut® will calculate the ideal mass tolerances for data extraction and scoring based on its extensive mass calibration. However, you can also specify your preferred tolerances for both MS1 and MS2 levels. You can choose amongst:

- Dynamic (default): determined by Spectronaut based on the extensive mass calibration. You can set a correction factor (default is no correction = 1)
- Relative: set a relative mass tolerance in ppm of the target ion mass
- Static: set a fix mass tolerance in Thompsons.

Calibration

Basic Settings	
MZ Extraction Strategy	Defines which MS peak within the m/z tolerance will be picked during mass calibration • **Maximum Intensity** (default): use the data point with highest intensity value within the calibration training tolerance. • **Nearest Peak**: Pick the peak closest to the theoretical m/z within the calibration training tolerance. This is useful in labeled experiments with extremely narrow mass differences (<50 ppm) between the channels.
Advanced Settings	
Allow source-specific iRT Calibration	Enable or disable source-specific iRT Calibration. If selected (default) Spectronaut will generate libraries containing as many iRT values as different sources exist in the dataset. (Box 7).
Precision iRT	If selected and available, Spectronaut will use a larger set of calibration peptides to perform an extensive calibration and improve iRT precision and accuracy. Exclude Deamidated Peptides (default and recommended): will exclude all peptides from the spectral library that contain deamidation from the calibration process. iRT ←→ RT Regression Type: • **Linear Regression**: will do a linear iRT calibration **Local (Non-linear) Regression**: this allows to calibrate non-linear gradients, but also to correct for possible variations of a linear gradient. If Spectronaut does not find enough data points, it will automatically change to linear calibration.

MS1 Mass Tolerance Strategy	Allows to determine MS1 tolerance for XIC extraction and scoring during calibration. One could choose from the following three options: • System Default – Spectronaut chooses MS1 tolerances specific for the different mass analyzers, namely: • Any Orbitrap +/- 40ppm • Any TOF (other than Bruker and Waters) +/- 40ppm • Bruker TOF +/- 40ppm • Waters TOF +/- 80ppm • Any Ion Trap +/- 0.5th • Relative – target mass tolerance is selected in relative units – ppm Static – a fixed mass tolerance is selected in Thomsons.
MS2 Mass Tolerance Strategy	• Allows to determine MS2 tolerance for XIC extraction and scoring during calibration. The option one can choose are the same as described above for MS1 Mass Tolerance Strategy.

Identification

Basic Settings	
Precursor PEP Cutoff	Specify the Posterior Error Probability (Precursor PEP) cutoff that should be considered identified. This primarily affects which precursors are quantifiable. When using either the global or run-wise imputation strategy, precursors that do not satisfy the PEP cutoff will be imputed.
Precursor Qvalue Cutoff and Protein Qvalue Cutoff	Choose your q-value (FDR) cutoff on precursor and protein level. Only those passing the cutoff will be considered identified and use for other subsequent processes. The protein Qvalue cutoff should be specified on both: experiment and run-wise level.

Single Hit Definition	Define what should be considered a protein single hit: stripped sequence, modified sequence, or peptide precursor ID. If only a single instance of the selected definition is identified across the entire experiment, the protein group will be marked as a single-hit protein group.
Exclude Single Hit Proteins	Discard protein groups identified with only one peptide hit (as defined above).
Advanced Settings	
Exclude Duplicate Assays	Spectronaut will keep only the best performing assay if a peptide is duplicated in the libraries.
Generate Decoys	• If unchecked, decoys have to be provided *a priori* in the library for Spectronaut to estimate the q-values (Qvalue). In such case the decoys need to be annotated in the column "**IsDecoy**" with the label "TRUE", and the targets need to be annotated with the label "FALSE". • If checked, you can specify the following: Decoy Method: defines how to generate the decoys. For details, please use the text hovers in the software. The default option is Mutated. Furthermore, since Spectronaut 14, the setting of Preferred Fragment Source was introduced. It could be chosen between two options: Template Fragments (will carry over the fragmentation pattern from the template peptide and just recalculate the masses based on the new sequence; recommended when other than HCD or CID fragmentation is used) and NN Predicted Fragments (neural network based strategy; will generate the ideal fragmentation pattern based on the newly generated decoy sequence). Decoy Limit Strategy: set the maximum number of decoys to be generated: • **Dynamic**: specify the number of decoys a fraction of the number of targets. • **Static**: choose a decoy limit as a fixed number of decoys. **None**: generate the same number of decoys as targets.
Machine Learning	• Per Run (default): calculates the discriminant scores (Cscores) and q-values (Qvalues) per run. • Across Experiment: makes a experiment-wise Cscore space. Can compromise the sensitivity.

Pvalue Estimator	Specify how you prefer the null distribution to be estimated to calculate the p-values: Kernel Density or Normal Distribution Estimator.

Quantification	
Basic Settings	
Precursor Filtering	Decide how to apply the q-value filter on the precursors in an experiment-wide manner to quantify protein groups: • **Identified (Qvalue)** (default): only those precursors passing the q-value cut-offs will be reported (considered as quantified) and, accordingly, used for statistical testing of differential abundance. Those cases not passing the cut-off will be tagged as "Filtered". By the default there will be no imputation of missing values. • **Identified in All Runs** (complete): the peptide precursor needs to pass the q-value threshold in all the samples to be reported. This is the most stringent filter and produces the smallest data matrix. • **Identified in % of Runs:** the peptide precursor needs to pass the Qvalue threshold in the user defined percentage of all experimental samples. For instance, if you set a 50th percentile cutoff (0.5), the peptide precursor needs to pass the Qvalue in 50% or more of your samples to be reported. In the samples where it was not identified below the significance threshold (default ≤ 0.01), the specified Imputing Strategy will be applied. Imputing strategy: the imputing strategy defines how to estimate the missing values (identifications not fulfilling the FDR threshold). • **Use Empirical Noise:** the best real picked signal will be reported. • **Global Imputing**: missing values are imputed based on a random sampling from a distribution of low abundant signals taken across the entire experiment. • **Run Wise Imputing**: missing values are imputed based on a random sampling from a distribution of low abundant signals taken within the corresponding run. This is useful for large scale experiments.
Proteotypicity Filter (only with automatic inference)	Choose whether you want to quantify only based on non-shared peptides, either at the level of protein (very stringent) or at the level of protein group.

Protein LFQ method	Specify how protein level label-free quantification should be performed. • **Automatic** will pick MaxLFQ for smaller experiments (<= 500 Runs) and Quant2.0 for experiments exceeding this cutoff. Such setting is recommended due to exponential run-time behavior of MaxLFQ algorithm. • **MaxLFQ** derives label-free quantities based on inter-run peptide ratios. **Quant 2.0 (SN standard)** is based on aggregation of minor group quantities based on user specified TopN and summation strategies
Quantity MS-Level	Choose which MS level you want to use to perform quantification: MS1 or MS2. MS2 level = sum of the quantities of the number of fragment ions per precursor, as specified in the spectral library.
Quantity type	Decide which feature of the peaks should be used for quantification: area under the curve within integration boundaries or apex peak height.

Cross Run Normalization	Apply label-free normalization to the whole dataset. Choose among the available Normalization Strategies: • **Automatic**: peaks between local and global normalization based on the number of runs (for n<500, the local normalization will be performed). • **Global Normalization**: choose to use median, average or geometric mean. • **Local Normalization**: based on the Local Regression Normalization described by Callister *et al.* 2006. **Normalization Filter Type** allows to limit the selection of peptides that are used for the normalization by specifying a filter: • **None** will use no specific filters for selecting peptides to use for normalization • **Library Name** will select peptides based on the name of the library • **FASTA Name** will select peptides based on the name of the associated parent FASTA file • **Modification Type** will allow selection of peptides for the normalization based on their modification state. **Row Selection**: choose which precursor profiles to use for normalization: • Automatic: 10000 precursor profiles, ranked by level of completeness. • Qvalue sparse (see explanation in the Data Filtering options). • Qvalue complete (see explanation in the Data Filtering options). Qvaule percentile (see explanation in the Data Filtering options).
<u>Advanced Settings</u>	
Interference Correction	Exclude fragment ions detected as interferences across all runs (Bilbao et al., 2015). If checked, set a minimum number of features to be kept at MS1 and MS2 levels in order for Spectronaut to still perform the quantification. Exclude All Multi-Channel Interferences: for labeled experiments, exclude the fragment ions that are shared between heavy and light channels and would be extracted from the same scans.

Major (Protein) Grouping	Specify what should be considered as a protein (Protein Group Id or Gene Id).
Minor (Peptide) Grouping	Specify what should be considered as a peptide (Stripped Sequence, Modified Sequence or Precursor).
Major Group Quantity	Specify how you want the minor groups (peptides) to be used to calculate the major group (proteins) quantities.
Major Group Top N	Use the best N minor group elements to calculate the major group quantities. The elements are ranked by evidence count and quantity.
Minor Group Quantity	Specify how you want the precursors to be used to calculate the minor groups (peptides) quantities.
Minor Group Top N	Use the best N precursors to calculate the minor group quantities. The precursors are ranked by evidence count and quantity.

PTM Workflow

Here you can specify settings related to the PTM analysis

Basic Settings	
PTM localization	Calculates a PTM localization probability for all variable modification site options. A specified probability cut-off can be applied (default is 0.75).
PTM Analysis	Performs a PTM focused data analysis: • **Hierarchical Clustering** specifies whether or not to run hierarchical clustering also on PTM sites. • **Multiplicity** will specify whether to differentiate singly from multiply modified peptides during the site collapse. If checked, doubly and triply modified peptides will generate separate collapsed site objects to be analyzed during regulation analysis. This only applies to modifications of the same type.

Advanced Settings	
PTM Analysis	**Flanking Region** specifies how many amino acids should be considered for the “PTM.FlankinRegion” report field. **PTM consolidation** specifies how to derive quantity from set of parent peptides carrying a particular modification on a given modification site: • **Sum** would summarize quantities of all parent peptides that are carrying particular modification on a given modification site. • **Linear model** would firstly impute missing values for each parent peptide based on quantities reported in other runs. Afterwards, it would summarize quantities of all parent peptides as above.

Workflow

Here you can specify if you are running a label-free analysis or a different kind of quantification

Basic Settings	
Method Evaluation	Will perform a separate Pulsar DIA search per condition in order to better compare different DIA methods within one experiment. This workflow is not meant for the quantitative experiments.
MS2 DeMultiplexing	Allows the processing of alternating shifted MS2 windows as presented in D. Amodei et al. 2019: • **Automatic** detects automatically of whether MS2 demultiplexing is required or not. • **Enabled** forces MS2 demultiplexing • **Disabled** forces no MS2 demultiplexing

Profiling Strategy	The profiling workflow allows the user to carry over the measured iRT of peptides that could be identified (q-value ≤ 0.01) in certain runs in order to fix integration boundaries in runs where an identification could not be achieved: **Template Correlation Profiling**: takes the best peptide signal in all runs as a template to find low abundant signals in the rest of the runs. **iRT Profiling**: takes the best peptide signal in all runs as a template and translates the empirical iRT to the integration boundaries of the low abundant signals in the rest of the runs. Profiling Row Selection: choose a row-wise q-value threshold for to choose which rows to profile (Minimum, Average or none) with a specified q-value threshold. Profiling Target Selection: specify which precursors should be readjusted (non-identified precursor or automatic).
Run Limit for directDIA Library	Defines number of DIA runs which are randomly selected for the library construction in directDIA. The software records which specific runs were selected. If this number is set to -1, Spectronaut will use all the DIA runs for construction of that library.
<u>Advanced Settings</u>	
Unify Peptide Peaks Strategy	Unify the peak picking across different charge states of the same modified peptide based on the highest scoring instance.

Protein Inference

Spectronaut is able to perform protein inference using the IDPicker algorithm (Zhang et al., 2007). Protein grouping will be well defined and protein group counts will be comparable across search engines and spectral libraries. Spectronaut also checks which peptides are proteotypic. The options are:

- **Automatic**: (default) When using a Spectronaut formatted library (.KIT) and all information is available, Spectronaut will use the protein inference parameters (the FASTA file used and the specified protein cleavage rules) to re-calculate the protein inference based on all identified peptides in the experimental DIA evidence.
- **From Search Engine**: will keep the protein inference as defined by the search engine within the provided spectral library. The protein entries will not be re-grouped based on the experimental DIA evidence.
- **From protein-db matching**: you can overwrite the existing grouping by choosing a FASTA sequence database and specifying the protein cleavage rules.

Post-Analysis

Basic Settings	
Differential Abundance Testing	Choose whether you want Spectronaut to perform differential abundance testing (paired or unpaired Student's t-test) or not.
Differential Abundance Grouping	Select what the biological unit you want your results to be based on as defined in the quantification settings: Major (proteins) or Minor (peptides) Group. **Smallest Quantity Unit**: the unit on which to perform statistical testing. The default unit used for paired and unpaired t-test is protein group.**Use All MS-Level Quantities**: regulation analysis with combined MS1 + MS2 statistical model (Huang et al., 2019). The publication is to reference the combined MS1 + MS2 model for statistical testing, not the statistical test itself.
Calculate Explained TIC	The explained TIC is the proportion of the TIC that can be associated to identified peaks. Choose if and how the relative explained TIC is calculated. • **None**: Will not calculate the explained TIC chart • **Extensive**: Will use an extensive feature detection method to calculate the most accurately for the explained TIC plot **Quick**: will use a simple and fast feature detector and correct the results based on a heuristic technique.
Calculate Sample Correlation Matrix	Choose whether you want to calculate the sample correlation matrix. If selected, a new plot will be available in the Post Analysis perspective, Analysis overview node, called Sample Correlation. For large experiments it can be very time consuming.
Hierarchical Clustering	Choose whether you want to cluster your samples and potential candidates.
Advanced Settings	

Differential Abundance Testing	• **Assume Equal Variance** specifies if the unpaired two sample t-test should assume equal variance between the test groups • **Group-Wise testing Correction**: Perform the multiple testing correction on groups of IDs within compared conditions. If not selected, the multiple testing correction will be performed on experiment level over all conditions. • **Log2 Ratio Candidate Filter** defines default log2 ratio filter applied to the candidates-grid in the post-analysis perspective. • **Confidence Candidate Filter** defines the default confidence type and cutoff to apply to the candidates-grif in the post-analysis perspective
Hierarchical Clustering	Defines further how Hierarchical Clustering should be performed.

Pipeline Mode

These settings are only relevant when running analyses from the Pipeline Perspective or from command line.

Specify if you want your analysis to be saved (with or without ion traces), which reports should be generated and saved, etc.

7.4 Appendix 4. Library Generation Settings

Configure the library generation settings. Most settings are described below. For further information, there are also helpful text hovers directly in the software.

Tolerances

Spectronaut® will, by default, calculate the ideal mass tolerances to generate the library. Spectronaut performs two calibration searches: based on the first-pass calibration (rough calibration), the ideal tolerance for the second-pass calibration is defined; based on the second-pass calibration (finer calibration), the ideal tolerance for the main search is defined. Spectronaut will do this under default settings (Dynamic).

However, Spectronaut allows you to set you preferred tolerances for the different MS instruments (Thermo Ion Trap, Thermo Orbitrap, TOF). Hence, for both the calibration search (second-pass, finer calibration), and the main search, you can define your tolerances:

- **Dynamic**: determined by Spectronaut based on the precedent search (default). You can set a correction factor for MS1 and MS2 levels (default is no correction = 1)
- **Relative**: set a relative mass tolerance in ppm for both MS1 and MS2 levels
- **Static**: set a fix mass tolerance in Thomson for both MS1 and MS2 levels

Identification

You can specify the search engine scoring type and thresholds

Pulsar	• Filter on PTM Localization: select this option when interested in including modified peptides only if they fulfill a PTM site localization probability (default 0.75). • Peptide FDR Cut off: Specify the FDR threshold on peptide level • Protein FDR Cut off: Specify the FDR threshold on protein level • PSM FDR Cut off: Specify the FDR threshold on PSM level
MaxQuant	• FDR Threshold: default set to 0.01
Proteome Discoverer	• Peptide Confidence Level: default set to High • Protein Confidence Level: default set to High • PSM Confidence Level: default set to High
ProteinPilot	• Peptide Confidence Level: default set to 99%

	• Protein Unused Score: default set to 0
Mascot	• Ion Score: default set to 0 • OneinXprobRnd: Ion score threshold value for e PSM being a random match based on 1 in p probability of random match

Protein Inference

You can have this option activated or deactivated. If you let Spectronaut do your proteins inference, you can refine your sequence settings further in these settings:

Enzyme/Cleavage Rules	Proteases used to *in silico* digest the proteins from the protein database(s). Defined in Databases → Cleavage Rules
Digest Type	Specific: both N- and C-terminus follow the specified digest rules Semi-specific: only of the termini follows the specified digest rules Unspecific: no digest rules
Toggle N-terminal M	Pre-processing of the protein database by toggling (processing both with and without) the protein N-terminal methionine (when there is one) to account for N-terminal methionine excision.

Spectral Library Filters

You have a number of options to filter the search engine results for library generation. There are filters at the level of fragment ion and at the level of precursor. The filters are quite self-explanatory. Please, use the hover text-tools if you need more information. Find below some of the most relevant.

Fragment Ions	Filter peptides not fulfilling the conditions specified regarding fragment ions. Find more details by hovering over the option in the software. You can specify your defined criteria for: • Amino acid length • Ion charge • Ion type (y, b) • m/z window • Relative intensity (%)

Precursors	Amino Acids: Filter peptides containing specified amino acids Best N fragments per Peptide: specify the range in number of fragment ions based on response Best N Peptides per Protein Group: keep only the N most abundant peptides per protein Channel Count: keep peptides that are observed in a range of channels FASTA matched: keep only peptide that are found in a user-specified FASTA sequence database and abide the digest rules. Only available if protein inference is selected. Missed Cleavage: keep peptides with N or less missed cleavages Modifications: Filter peptides according to modifications. Find more details by hovering over the option in the software Peptide Charge: keep peptides with Proteotypicity: If a peptide has a unique protein mapping (matched to a single FASTA entry from the sequence database).

iRT Calibration

Set your preferences for iRT calibration:

Basic Settings	
Calibrate from empirical RT	Generate iRT values from empirical RTs during library import
iRT Reference Strategy	Define how the reference iRT is derived for iRT calibration: • Deep Learning Assisted iRT Regression. Use the new Deep Learning algorithm to generate the iRT reference set. This is useful when working with non-model organisms hardly covered in Spectronaut's internal empirical iRT reference dataset. • Empirical iRT Database. Use Spectronaut's internal empirical iRT reference database of more than 100.000 iRT reference peptides from multiple sources.

	Use RT as iRT. No iRT calibration will be performed. It should only be used if the peptide separation method is very stable, homogeneous and non-standard, such as capillary zone electrophoresis (CZE).
Advanced Settings	
Minimum Rsquare	Choose how strict you want to be to accept the iRT calibration of your data by specifying the minimum coefficient of determination (R^2) allowed.

Workflow

Fragment Ion Selection Strategy: defines the strategy to be used for selecting the top N fragment ions per peptide precursor

- **Intensity Based**: Prioritizes fragment ions by their intensity in the consensus spectra
- **Evidence Based**: Prioritizes fragment ions by how often they have been observed in the experiment for the same precursor
- **Maximize sequence coverage**: Groups fragment ions by type and position and ranks them based on the best fragment ion per group (either by intensity of by evidence) in an iterative manner.

In-Silico Generate Missing Channels: if generating a library for an isotopically labeled sample, you can activate this option to *in-silico* generate an assay for a missing label pair. For example, if you have a sample of only spiked-in heavy aqua peptides, it will make a light/heavy library where the label-free channel is generated *in-silico* based on the heavy channel and with a consistent fragment ion selection. This option works with n-channel search results as long as you are working with isotopic labels. The workflow options are:

Spike in workflow: will create a light channel for all heavy (SIS) peptides that are identified without a light counterpart.

Labeled workflow: will detect the labeling setup of an experiment and add the channels that are missing for a given peptide

Inverted spike-in: will create a heavy channel for all light peptides that are identified.

Use DNN Ion Mobility decides if ion mobility should be predicted based on deep neural network (DNN) for library generation.

- **Auto** will always predict Ion Mobility during library generation. Only if empirical ion mobility value is not available for the peptide, a predicted value will be used.
- **Always use predicted Ion Mobility** – a library will contain predicted ion mobility values, regardless if empirical information is available or not.
- **Never predict Ion Mobility** – ion mobility prediction will not be performed. Library will contain only empirical ion mobility values (if available).

7.5 Appendix 5. Analysis Perspective Plots

Run Node Plots

iRT Calibration Chart

This chart shows the status of run calibration. Spectronaut® supports non-linear gradients using a refined calibration based on the initial calibration and detailed information in the user library.

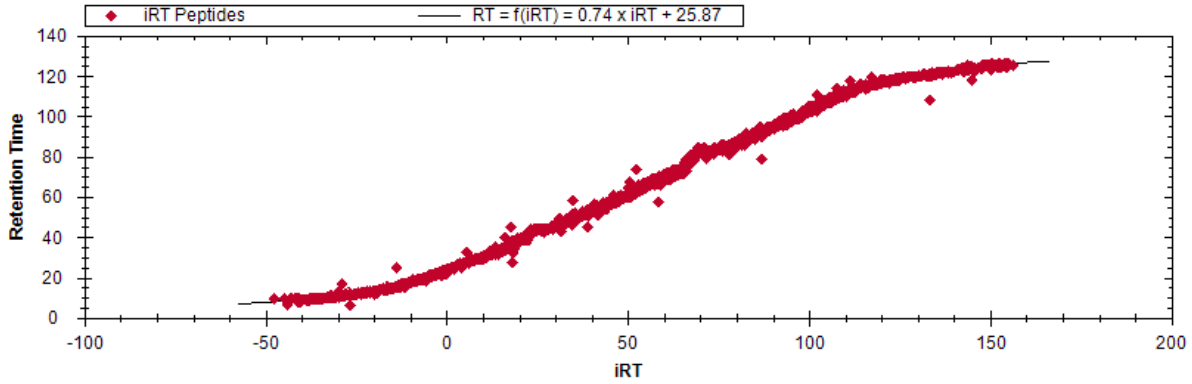


Figure 49. iRT Calibration Chart showing the non-linear transformation from library iRT to actual predicted retention times. The chart is shown after extended non-linear calibration has been performed. The extended calibration allows you to correct even small gradient fine-structure fluctuations to get the most accurate retention time prediction for your library.

XIC Extraction Width Chart

Based on extensive calibration using the respective calibration Kit and information about the spectral library used, Spectronaut can determine the optimal XIC extraction width for your experiment. This extraction width is calculated dynamically over the whole gradient and allows Spectronaut to automatically adapt to areas with lower retention time prediction accuracy. The window width can be influenced with a correction factor in the Settings perspective (Settings → Analysis → Peak Detection → Correction Factor).

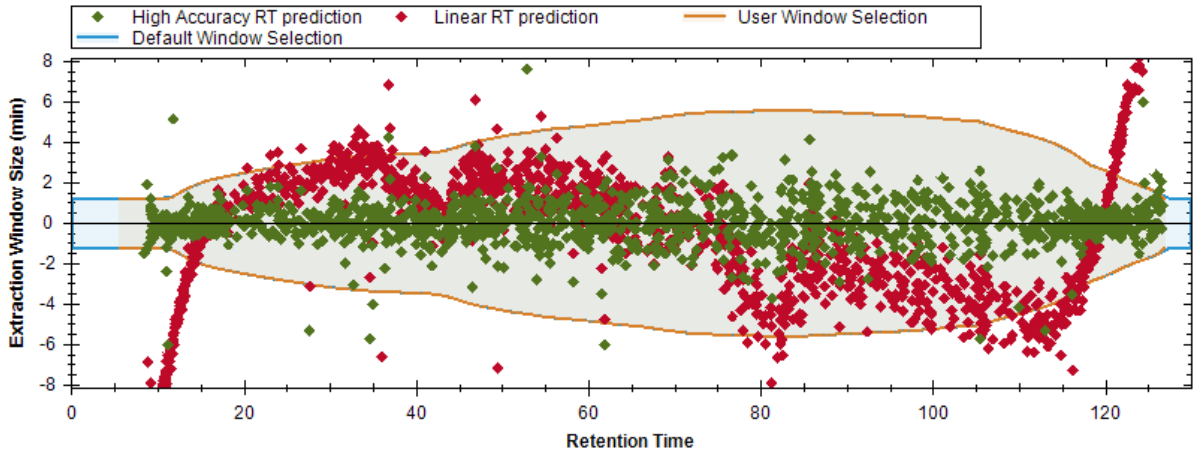


Figure 50. The XIC Extraction Width plot gives insight into your gradient stability and the overall accuracy of your library's iRT values. The blue and the orange lines show the window selections as suggested by Spectronaut (blue) and as set by the user (orange → correction factor in the settings). These extraction window widths (y-axis) change over time (x-axis) based on gradient stability and iRT accuracy. The red dot's show your libraries iRT accuracy assuming a linear iRT to RT transformation. The green dots show the iRT accuracy using the extended non-linear iRT to RT transformation. The later one is used for the actual analysis.

Ion Mobility Calibration

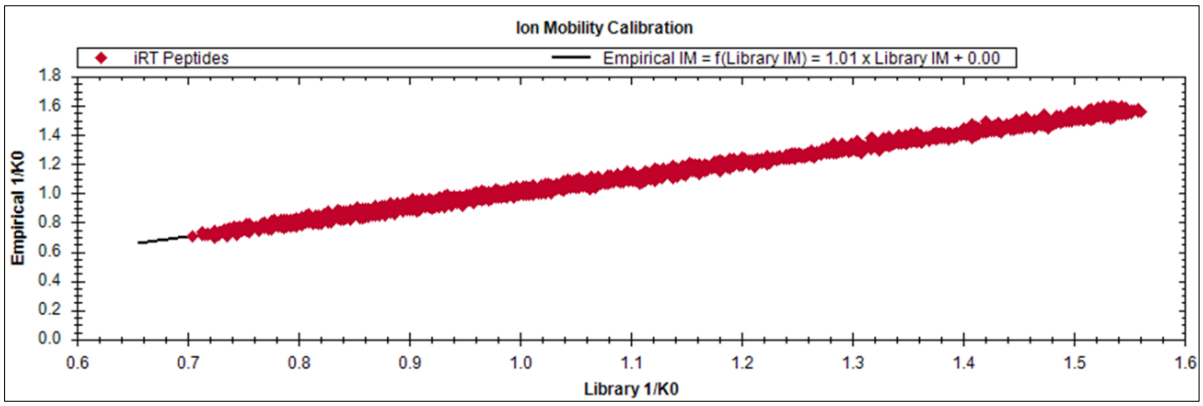


Figure 51. The Ion Mobility Calibration plot shows the empirical ion mobility as a function of the ion mobility values in the library.

Ion Mobility Extraction Width

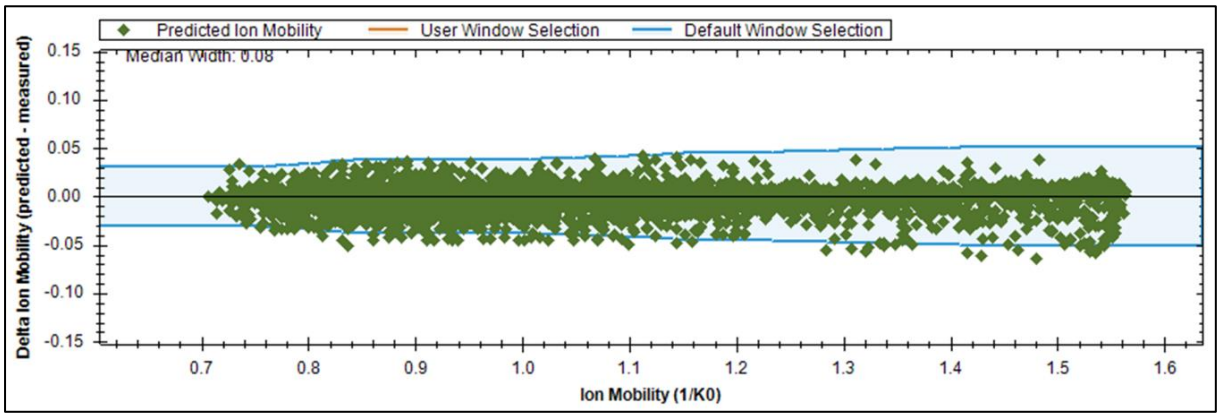


Figure 52. The Ion Mobility Extraction Width plot shows the difference between predicted and measured ion mobility for each ion mobility value as well as the applied extraction window.

MS1 TIC Chromatogram & Base Peak Chromatogram

The MS1 TIC Chromatogram plot shows the signal of all ions in function of RT. The Base Peak Chromatogram shows the signal of the most intense precursor ion at certain RT.

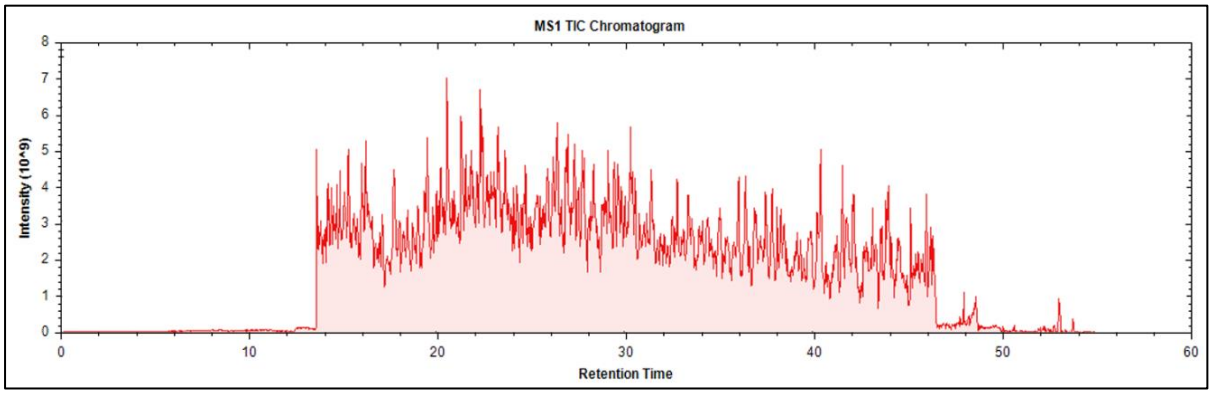


Figure 53. The MS1 TIC chromatogram shows the total ion current of a run, giving insight into the amount of sample injected.

Analysis Log

The analysis log contains all the information pertaining to the analysis of your whole experiment. In the event of errors one can consult the analysis log for detailed information of what went wrong.

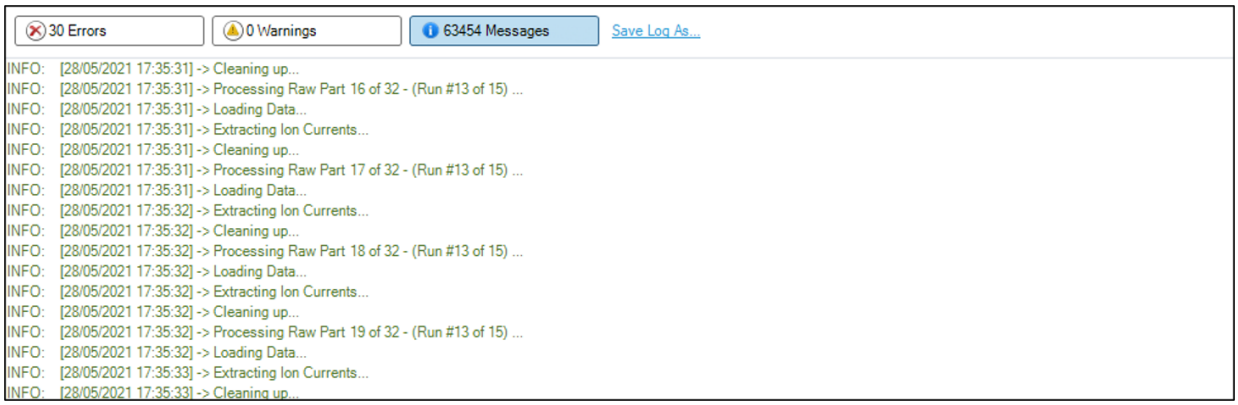


Figure 54. The analysis log with detailed information about the analysis processes in Spectronaut.

DIA Acquisition Method Overview

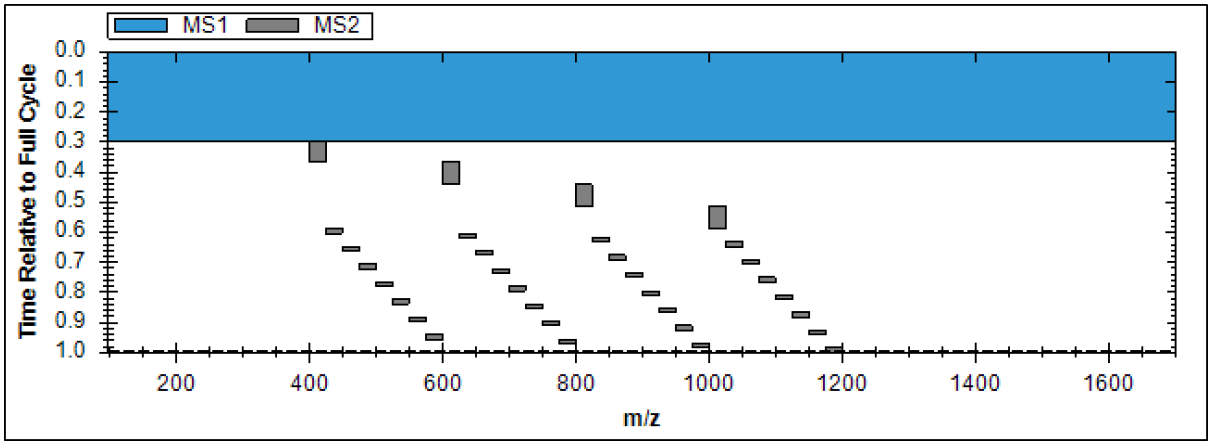


Figure 55. The DIA Acquisition Method Overview shows details on how the DIA acquisition method was constructed. It is a convenient tool for inspection of the correctness of the used method and serves the purposes of acquisition method development and optimization.

Ion Mobility Overview

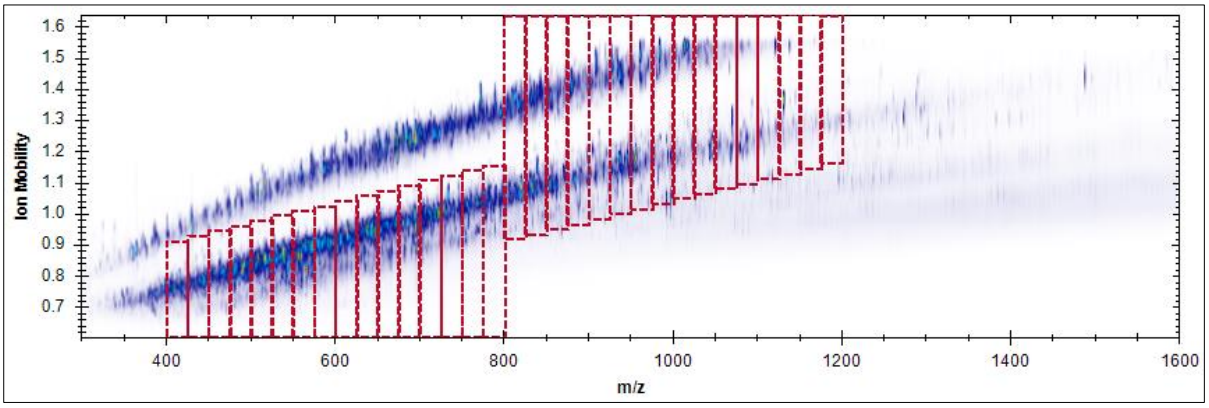


Figure 56. The Ion Mobility Overview plot shows the ion mobility dimension in function of m/z values. The red lines show consecutive DIA boxes. Their overlay with registered MS1 signal could be used for DIA method optimization and validation.

Mass Error Histograms

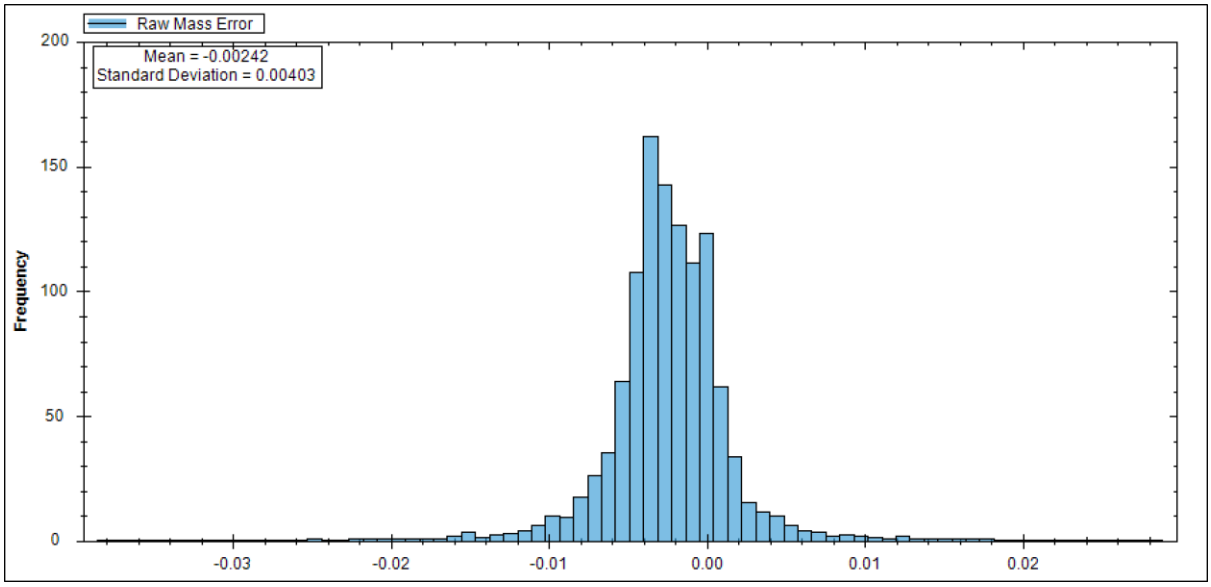


Figure 57. The mass error histogram at MS2 and MS1 levels can be used for data quality control and mass spectrometer behavior monitoring.

Peptide and Fragment Plots

MS2 XIC

The default plot on Elution Group (EG or peptide), Fragment Group (FG or peptide precursor) and Fragment (F or fragment ion) level. This plot shows the extracted ion current chromatogram of the selected peptide. The plot contains the XICs for all fragments present in the library. Additionally, the expected retention time is marked (black dotted line) and the currently selected peak for this peptide (green area). On the Elution Group and Fragment Group level the coloring of the fragments is based on its expected relative intensity ranging from red (expected most intense fragment) to blue (expected least intense fragment). You can directly change integration boundaries in the plot.

On fragment level, this plot only shows the selected fragment in color and all others in gray.

There are a number of options available upon right-clicking in the plot such as switching the y-axis to log scale, toggling accept/reject the peak, changing x-axis scale to iRT, showing normalized intensities and showing the XIC chromatogram for the whole gradient.

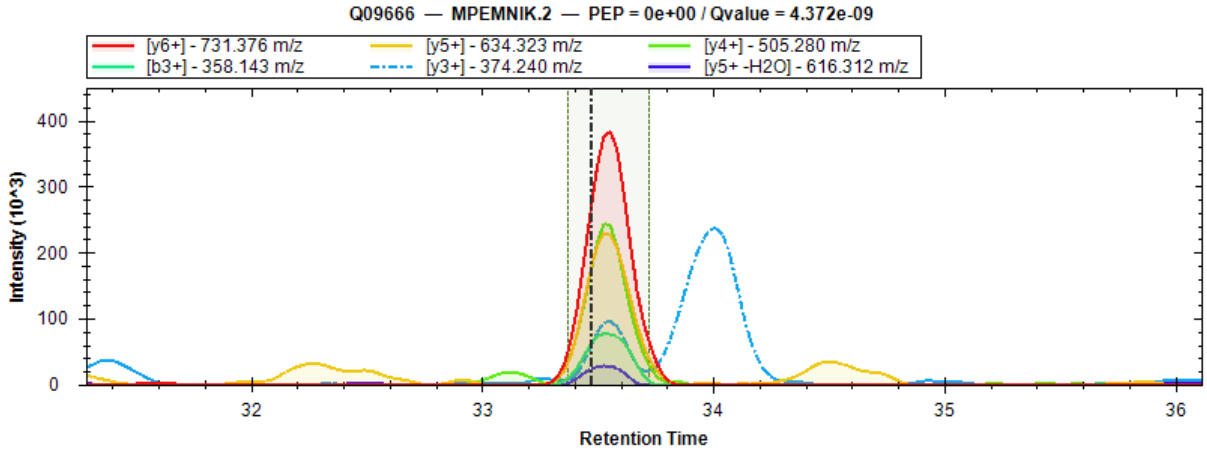


Figure 58. XIC chromatogram for the peptide MPEMNIK++. The color coding of the fragments indicates an overall good correlation to the expected fragment intensities. The dotted blue line indicates a potential interference that was detected by Spectronaut and automatically removed for relative quantitation.

MS2 XIC Sum

The XIC Sum chromatogram chart shows the selected peptides quantitative information. The XIC shown is the sum of all fragment XICs that qualified for quantitation. All fragments that were excluded due to interfering signals are not used to calculate the sum XIC chromatogram.

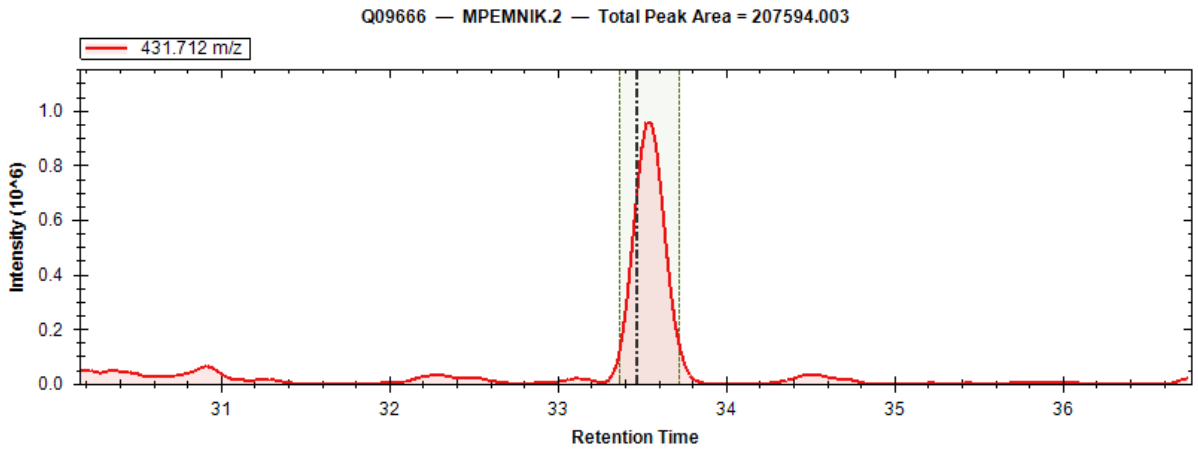


Figure 59. The MS2 XIC Sum plot of the peptide MPEMNIK++. The interference highlighted in the MS2 XIC plot (Figure 58) has already been removed from the sum XIC chromatogram.

MS2 Intensity Correlation

This plot shows detailed information about the correlation of predicted and measured relative fragment ion intensity. The predicted values in red correspond to the relative intensities provided by the spectral library. The black lines correspond to the relative measured intensity of each fragment ion. Fragment ions with potential interferences are displayed as dotted lines.

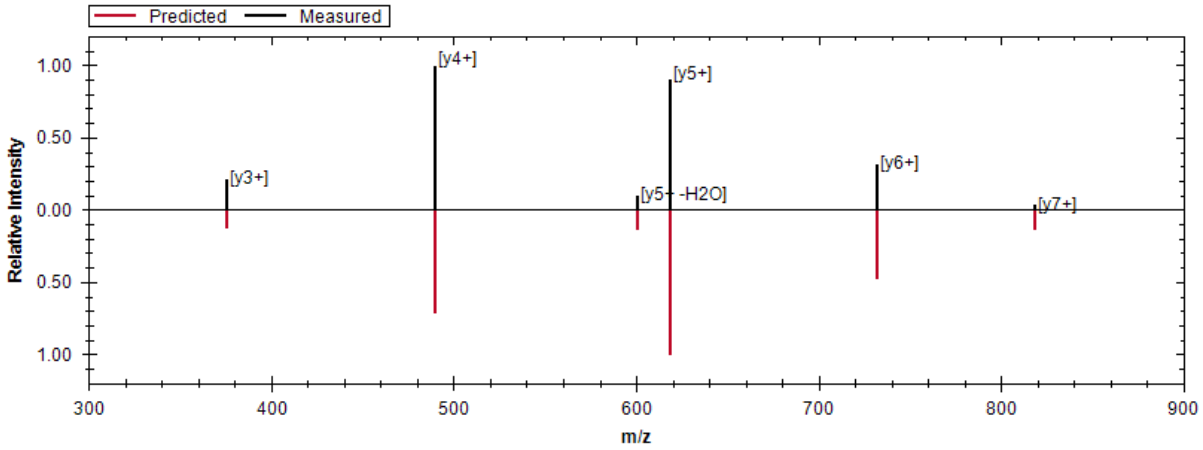


Figure 60. The MS2 Intensity Correlation plot for a given peptide precursor. The plot indicates a very good correlation between the expected relative intensities (red) and the measured intensities of the library fragments.

MS1 Isotope Envelope XIC

This plot shows the monoisotopic precursor plus its most abundant isotopic forms as an XIC chromatogram. The XIC chromatograms on MS1 and MS2 level are expected to have identical apex retention times and elution shapes. As with the MS2 XIC chromatogram plot, the MS1 XIC chromatogram is also color coded to reflect the predicted relative intensities. A color coding from red (highest) to blue (lowest) indicates high correlation with the predicted abundance.

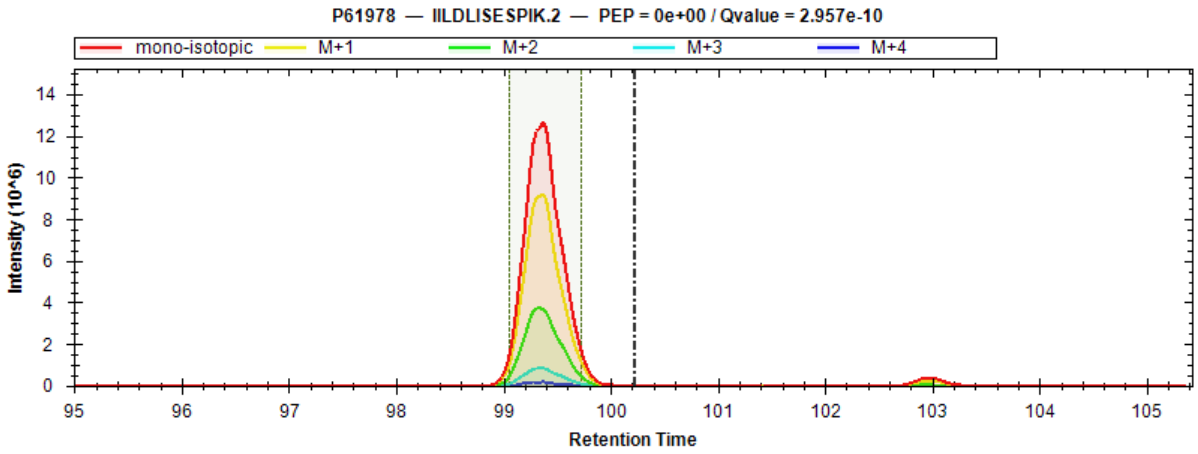


Figure 61. The MS1 Isotope Envelope XIC chromatogram for the precursor IILDLISESPIK++ as extracted by Spectronaut. The color pattern again indicates a high correlation to the expected relative abundances.

MS2 Isotope Envelope Correlation

Similar to the MS2 Intensity Correlation plot, this plot highlights the correlation between the expected and the observed fragment isotope patterns. A high correlation between the measured (black) and the predicted (red) abundances signals high confidence in the identification and quantification.

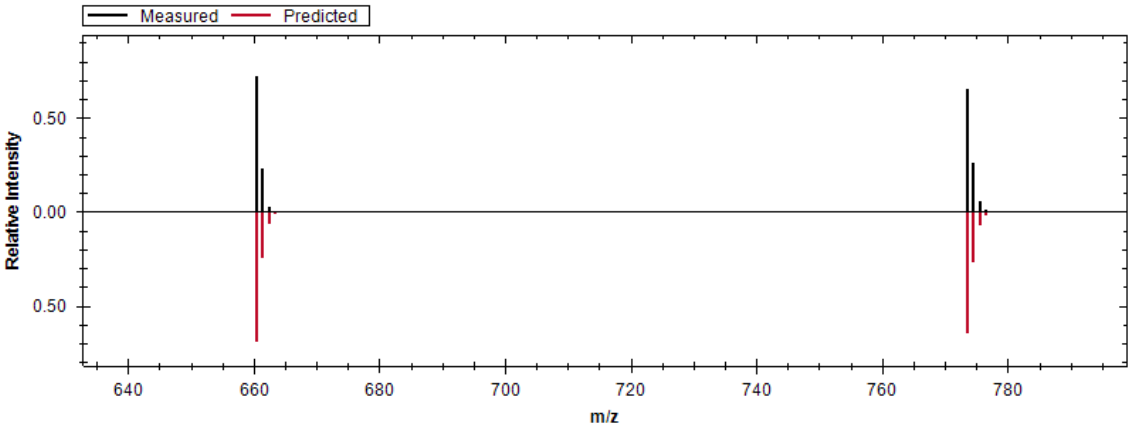


Figure 62. The envelope correlation plot for two fragments with very high correlation with respect to the predicted abundance. The measured intensities (black) are almost perfect mirror images of the predicted (red) intensities.

PTM Localization Plot

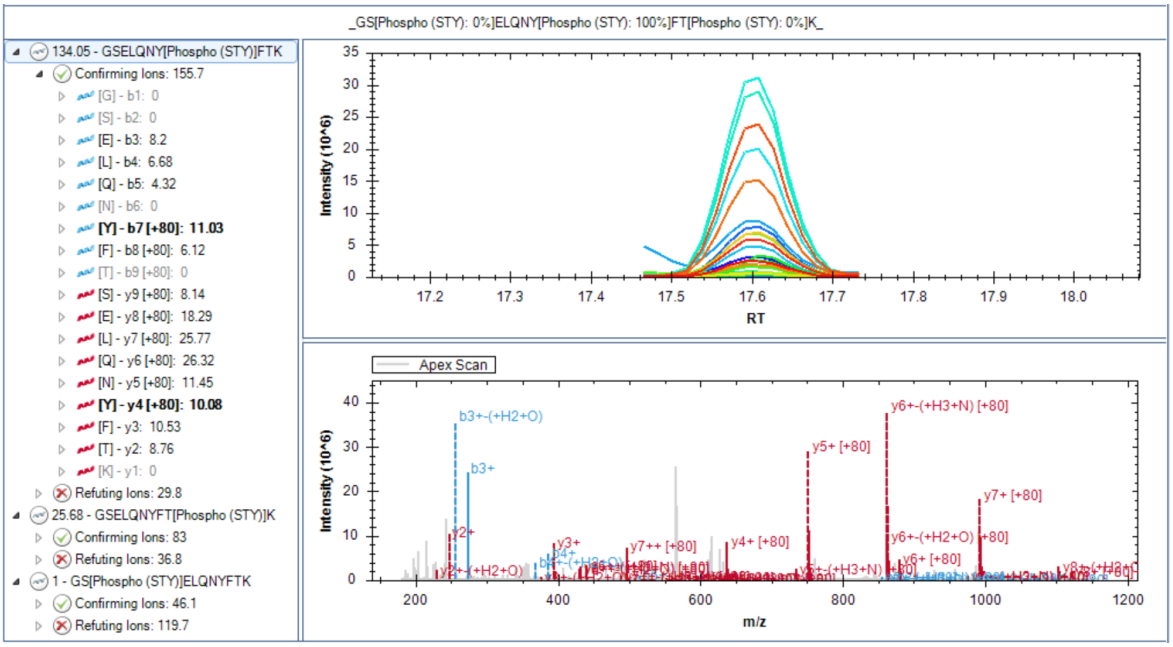


Figure 63. The new PTM localization plot shows the different possible modified versions of a peptide assay, depicting the corresponding scores for each of the fragment ions, either confirming the given site probabilities, or refuting it.

MS1 XIC Alignment and MS2 XIC Alignment Plots

The MS1 and MS2 XIC Alignment plots allow visualization of the extracted ion chromatograms of a peptide across runs. This visualization allows you to easily and quickly browse through thousands of XICs in a glance. By right-clicking on the panel you can also sort the grid by experiment conditions and replicates - conditions are sorted in rows and replicates in columns. The name above the XIC plot is a clickable link, which redirects you to that specific precursor in a run.

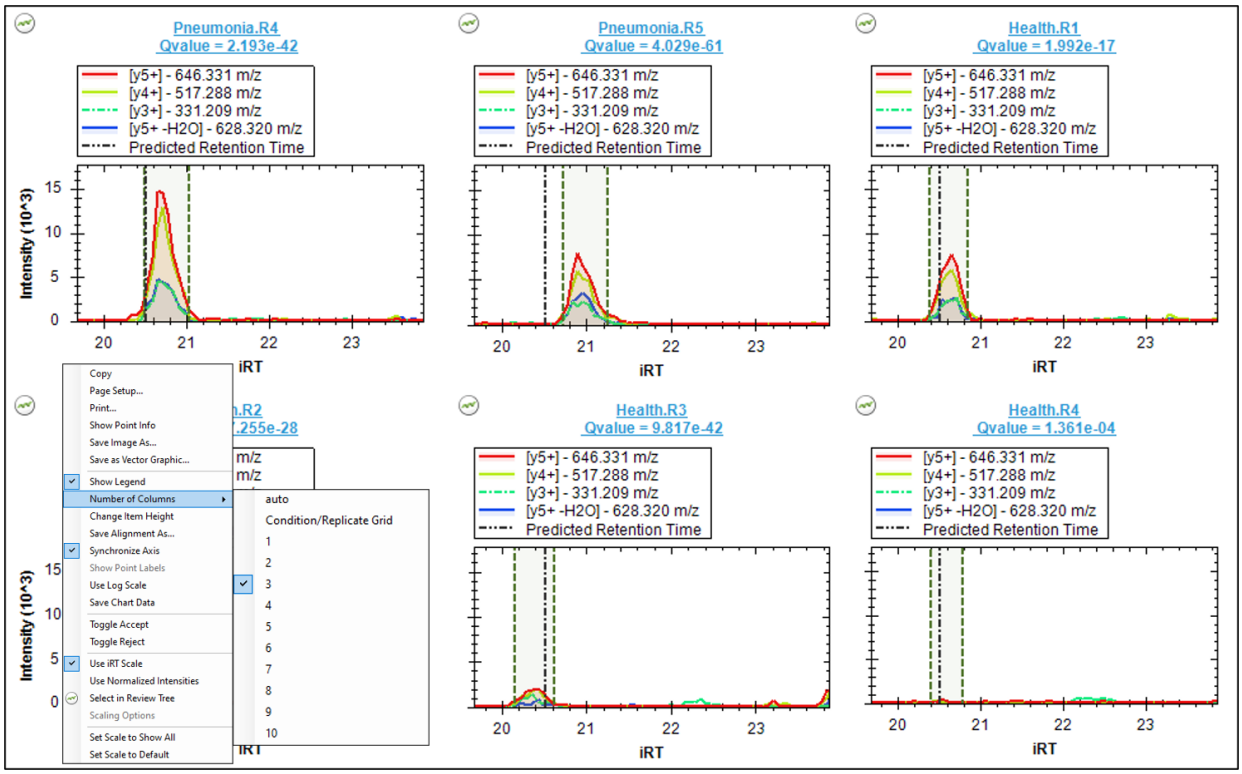


Figure 64. The MS2 XIC Alignment across runs. The x-axis is automatically changed to iRT to reduce chromatographic variance. The axis can be changed to retention time by right-clicking on the plot and un-selecting the "Use iRT Scale" option.

XIC graph

This plot combines the MS2 XIC and the MS1 Isotope Envelope XIC together and across all your runs. This is a nice way of looking how well MS2 and MS1 peaks correlates and how reproducibly thy behave across runs.

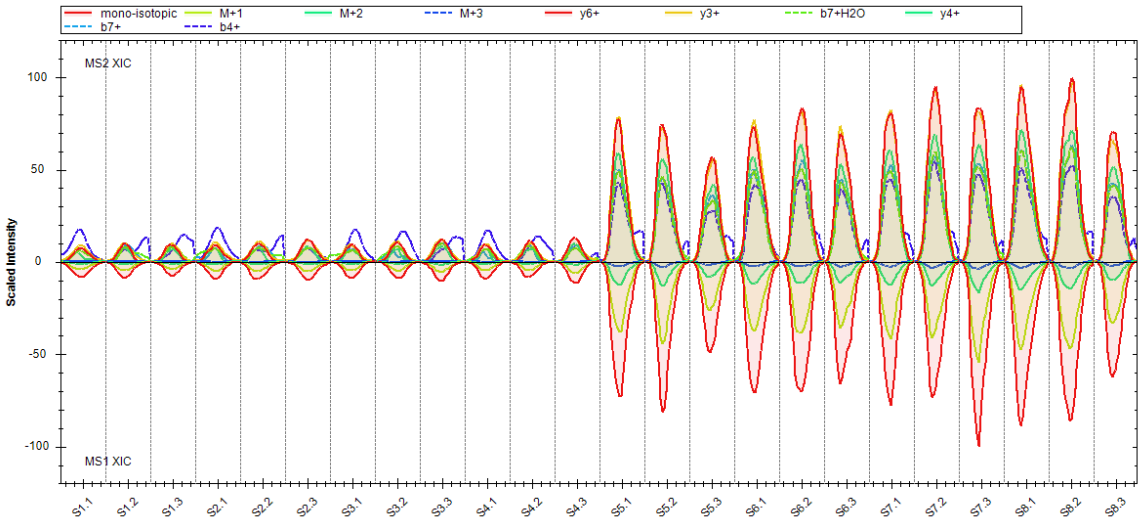


Figure 65. The XIC graph for one peptide over the 24 run files analyzed. The upper row shows the MS2 XIC. The dashed blue line indicates a potential interference. The lower row shows the MS1 Isotope Envelope XIC.

iRT XIC Sum Overlay

This plot allows you to show all sum XIC chromatograms of your selected peptide from all runs as an overlay plot. The x-axis scale is by default in iRT but on right-click can be changed to actual retention time.

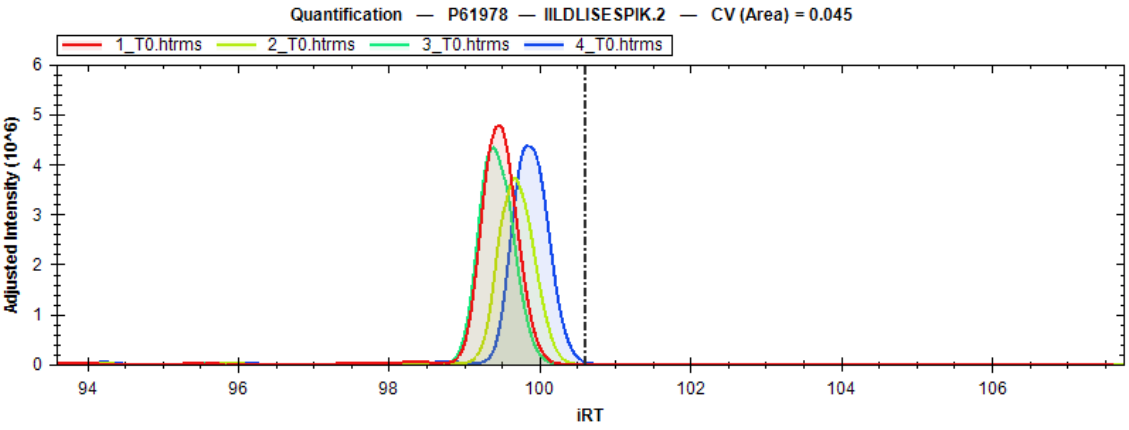


Figure 66. The iRT XIC Sum Overlay chart for the peptide IILDLISESPIK++. The 4 XICs correspond to the sum XIC of one peptide in the 4 different runs loaded for this experiment. The title of the plot additionally shows the peptides coefficient of variation, in this case 4.5%.

Ion Mobilogram

For data that use ion mobility as a fourth dimension for ion separation, such as dia-PASEF data, the mobilogram is available. The mobilogram allows for manual peak integration in ion

mobility dimension. By reintegrating the IM dimension, the RT XIC also changes since it is built from what is contained in the IM extraction window. If you want to reintegrate a peak, you can adjust the extraction boundaries and this will automatically affect the RT XIC.

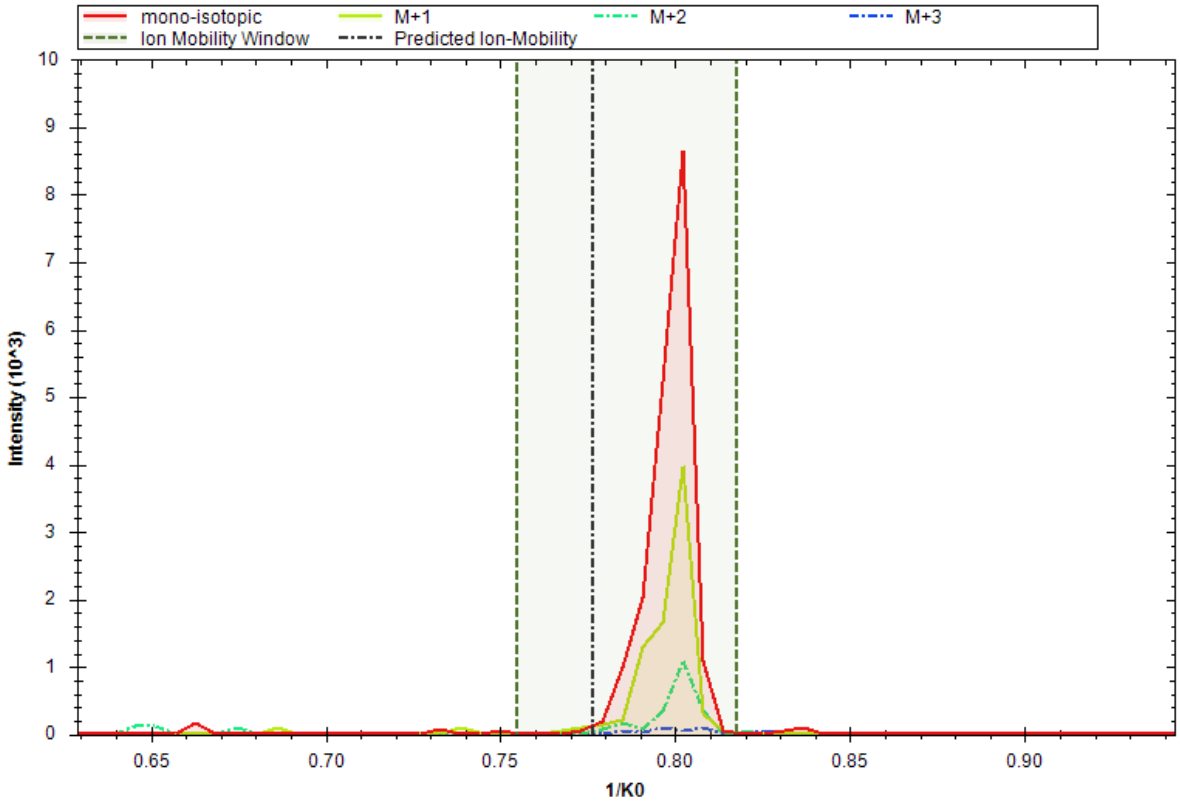


Figure 67.The ion mobilogram for visualization and peak re-integration in the ion mobility dimension.

MS2 Intensity Alignment

Similar to the MS2 XIC Alignment, this plot gives detailed information about the signal stability for one peptide across several runs. The different bars show the relative abundance of each fragment ion across multiple runs. Using this plot, one can quickly identify an inconsistent signal by the change in the color pattern. Right-click on the plot and un-selecting "Normalize" to show intensities on absolute scale.

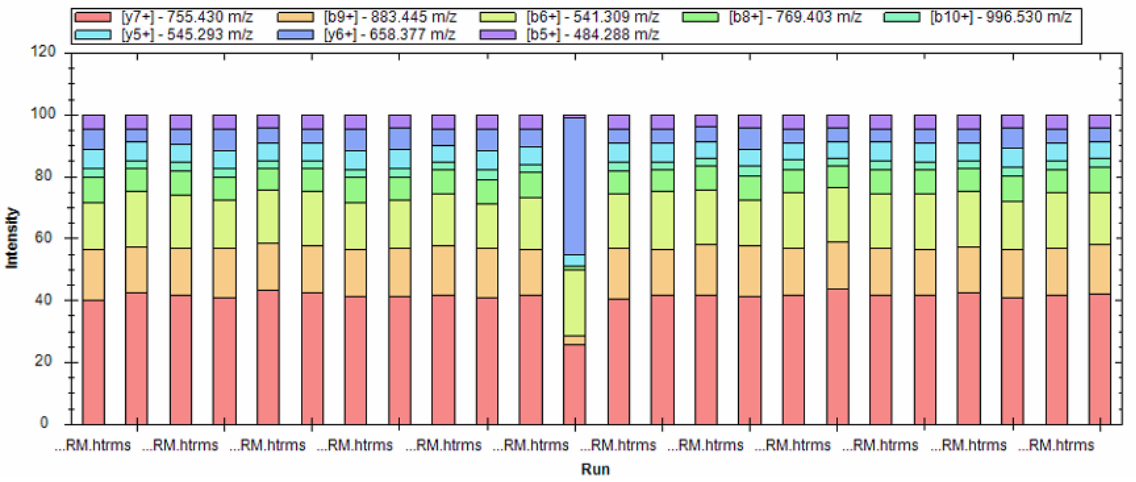


Figure 68. The MS2 Intensity alignment for a peptide containing 8 fragment ions. Each fragment ions relative intensity compared to the total peak height is indicated using a differently colored bar. The peptide was targeted in 24 runs. An inconsistent signal can be easily identified due to the sudden change in the color pattern.

Cross Run RT Accuracy

Similar to the MS2 Intensity Alignment chart, the Cross Run RT Accuracy plot allows one to quickly validate the peak picking across several LC-MS runs. The x-axis shows colored bars that correspond to the peptide in different runs. On the y-axis one can see the retention time in iRT.

The height of each box corresponds to the peak width at the start and the end iRT according to the y-axis. The line through the middle of the box shows the apex retention time in iRT while the blue colored boxes in the back show the total XIC extraction width. The colors of the bars again correspond to the relative intensities of the measured fragment ions. The bar with the green background is the currently selected node. The black, dotted, horizontal line corresponds to the expected retention time in iRT. You can hop to any other node by clicking on the colored bar.

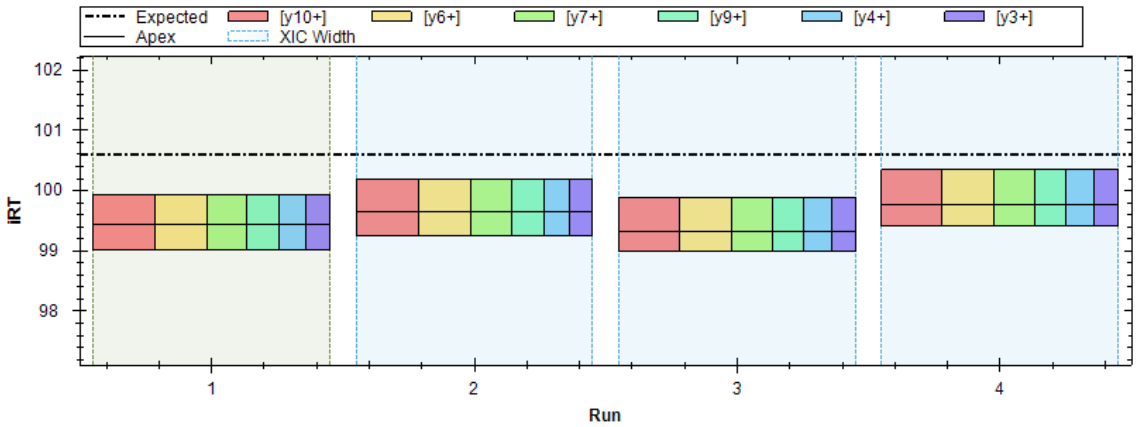


Figure 69. The Cross Run RT Accuracy plot for a peptide measured in 4 different runs. The multi-colored bars correspond to the detected peak with the colors encoding the relative fragment intensities and the upper and lower boundary of the bar corresponding to the peaks start and end retention time in iRT.

MS1 Spectrum at Apex

This plot shows the MS1 signal at apex retention time for a given peptide. The chart automatically zooms in on the isotopic envelope and labels the different peaks accordingly. The red bars indicate the expected relative intensities of the different isotopic forms of the precursor.

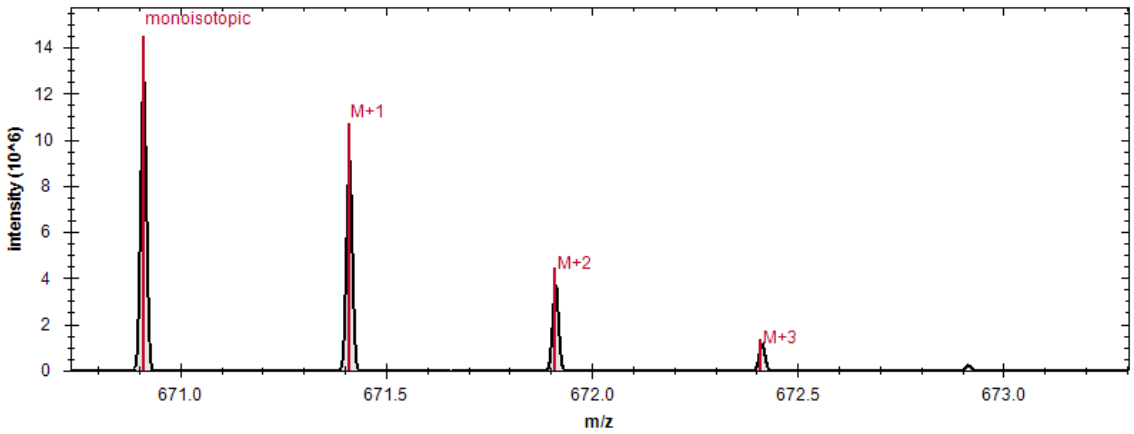


Figure 70. MS1 spectrum at apex for the peptide IILDLISESPIK++ showing the monoisotopic precursor plus its first 3 isotopic forms.

MS2 Spectrum at Apex

Similar to the MS1 Spectrum at Apex, this chart shows the full recorded MS2 spectrum corresponding to a given peptide's LC-peak apex. By default, the plot only highlights the fragments as provided by the library. Additionally, other fragments can be calculated and annotated by right-clicking on the plot and selecting "Show All Theoretical Fragments". You can also use the "Set Fragment Filter" window to open up the range of fragments and show, for instance, a-series ions.

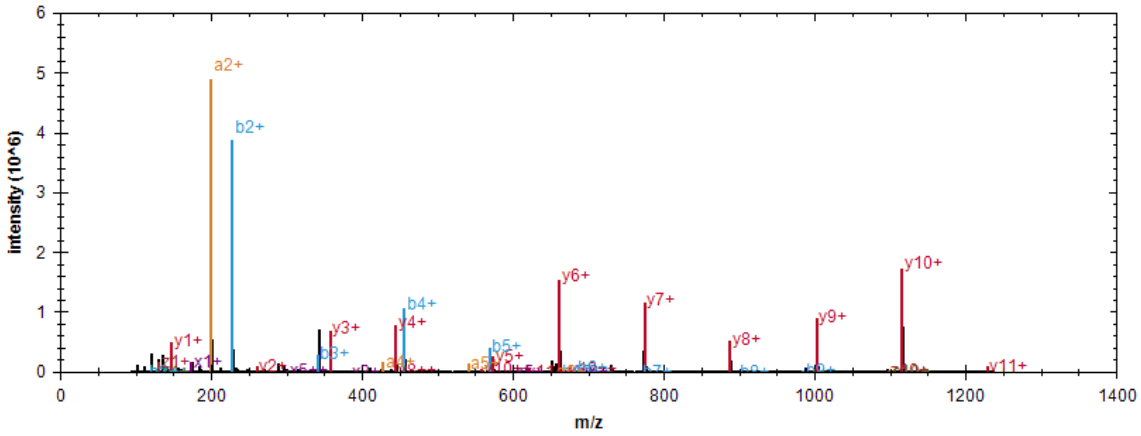


Figure 71. MS2 Spectrum at Apex for the peptide ILLDLISESPIK++. The option "Show All Theoretical Fragments" was turned on and a-ions were selected in the fragment filter editor.

Peptide Data Match (PDM) Plot

The PDM plot is newly introduced with Spectronaut 13. This plot shows the MS features for a precursor and how well it matches the expected values. The main area of the plot (on the right side) shows the full recorded MS2 spectrum corresponding to the peptide LC-peak apex. Highlighted, you can see expected fragments for that peptide (full line denotes the calibrated m/z value and the dashed line denotes the theoretical m/z value of a specific ion). At the bottom of this plot, the mass errors of each of the highlighted fragments are shown. On the left side, the MS1 spectrum at Apex plot is depicted (similar to the MS1 Spectrum at Apex plot described before, the full line denotes the calibrated m/z value and the dashed line denotes the theoretical m/z value of a specific ion). Finally, at the top-left side, you can choose among several elements to show on the MS1 and MS2 plots. Some of the options are:

- Label the fragment peaks with the predicted ion name or with the m/z of the peak
- Show the mass error in ppm or in Thompson
- Choose which fragments to highlight on the MS2 spectrum: the ones used for the assay (Library Fragments), the ones detected in the library but not selected (Extended Library Fragments), or All Theoretical. You can also filter by fragment length, by type and by neutral loss.

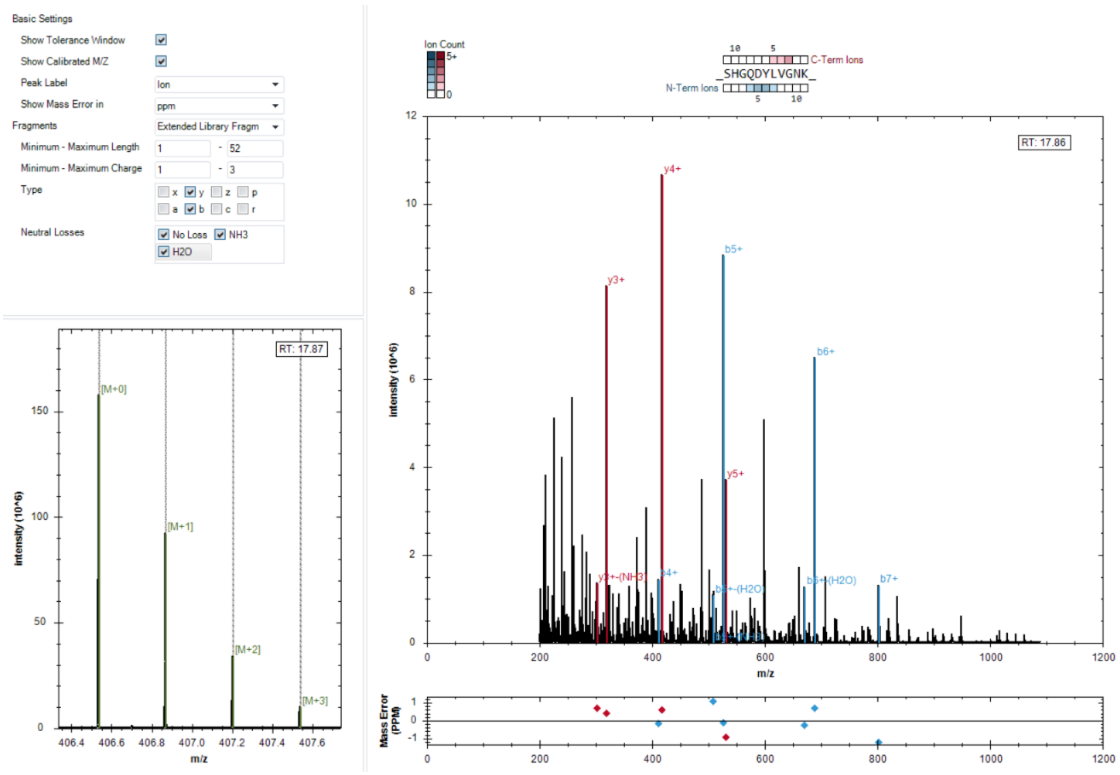


Figure 72. Peptide Data Match (PDM) plot for the precursor SHGQDYLVGNK+++. At the top-right panel you can select what to highlight on the MS1 and MS2 spectrum plots. In this case, all fragments in the library

fulfilling the filter criteria where shown. In blue you can see the b-ions: in red, the y-ions. The mass errors, in ppm, are shown for each fragment. The MS1 spectrum is shown at the bottom-left.

Protein Coverage

Spectronaut gives you a detailed overview of a protein's coverage within your analysis. The protein coverage plot shows you all the peptides of a protein that were targeted within your current experiment. The color coding by default indicates the confidence level for each peptide. This can be toggled on right-click to show the proteotypicity status for each peptide. This option is only available if protein inference was enabled during the DIA analysis. By choosing the Digest Specificity option, you can see if the identified peptides were digested specifically. For PTM workflows, you can also visualize PTM site annotation. Only confidently localized PTMs are highlighted in the plot. If you click on one of the peptides in the plot, you can additionally visualize its XIC.

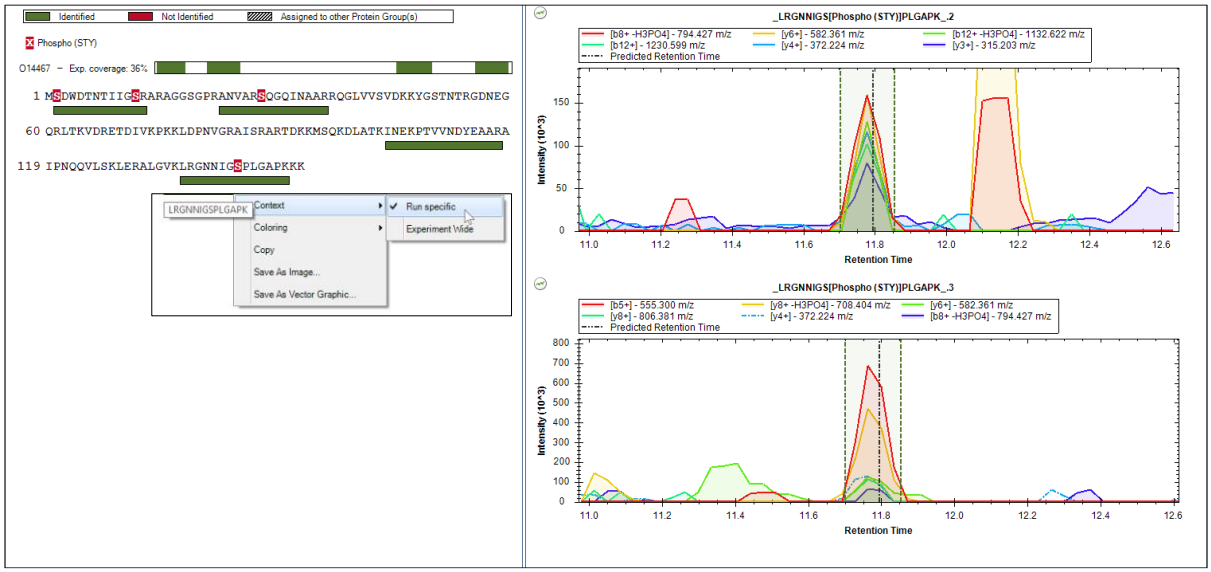


Figure 73. Protein coverage plot. The protein sequence coverage can be displayed at run- and experiment level. Right-clicking on a peptide also allows to see proteotypicity or digest specificity.

Condition Box Plot

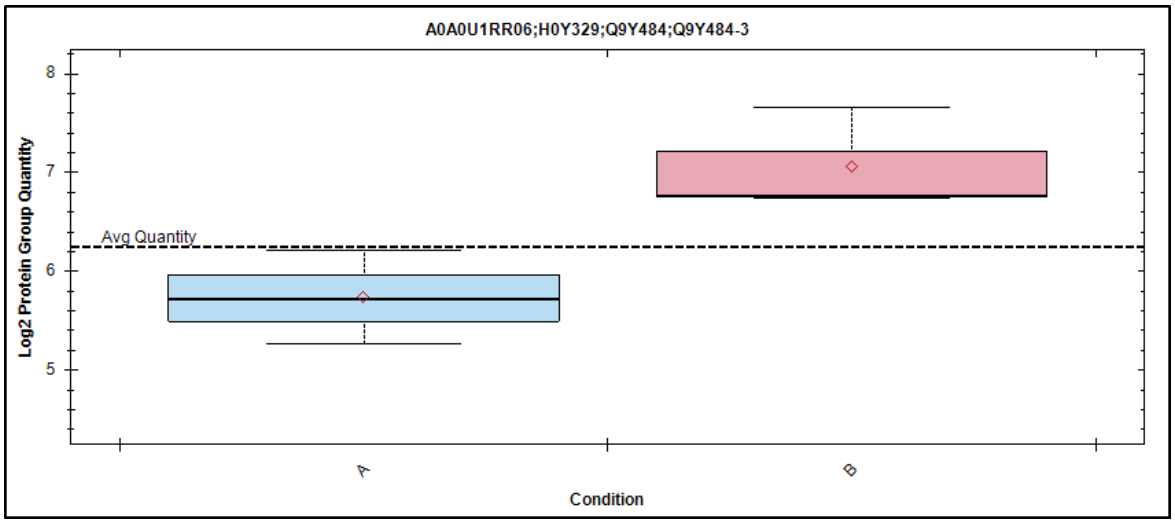


Figure 74. The Condition Box Plot in the Grid View shows the conditions average log2-transformed quantities. The color code is blue for lower and red for higher abundances. By right-clicking on the plot, the Base Quantity can be set to precursor, peptide, or protein group.

7.6 Appendix 6. Experiment Tab Options

In the Analysis Perspective, right-click on the experiment tab. A context menu will open with several functionalities to apply to the analysis (**Figure 22**). The most common actions are available in intuitive icons displayed below the experiment tab.

Add runs and Remove runs	Add or remove runs from the analysis. For the changes to take effect, you need to refresh the Post-Analysis.
Map missing runs	If Spectronaut® lost the link with the run files, you can map them back. If your analysis contains XICs, this is not needed.
Recalculate Qvalues	If you manually change the peak picking, by selecting a different peak of changing the integration boundaries, the q-values need to be re-calculated.
Refresh Post Analysis	If you do any modification on the analysis, such as manually modify a peak or add new runs, the Post-Analysis has to be refreshed.
Re-extract all XICs	If you saved your experiment without XICs, it is recommended that you re-extract them from the run files for a better performance.
Export Experiment Settings	Export a report with the settings you used to run your analysis
Order Run by	Choose a criterion to order your runs
Group by	You can group your data tree under different criteria. Default and recommended grouping is by protein group. If you have several workflows in your analysis (e.g., label-free and spike-in), Group by Workflow is very useful.
Reset All Peaks	Revert the manually modified peak picking back to the automatic one. The q-values need to be recalculated
Commit Library Changes	If you refined your library in the Analysis Perspective, the changes will not make effect until your commit the changes (see Section 3.4.1.11).
Save as and Save	Spectronaut will not save your analysis automatically. You need to actively save your experiment as. Choose to save with or without XICs (see Section 3.4.1.9).

Export All… (Ctrl + R)	With this function, you can batch export many reports of your choice relevant to your analysis.
Settings	This tool allows to explore and **change** many settings of your analysis: • Identification, Quantification, Protein Inference and Post Analysis. • Different FASTA files can be loaded here, as well as GO annotation files. • You can also change the conditions set-up • Finally, in the Summary tab, you can change the name of your analysis Changing these settings will let you recalculate your analysis, which is significantly less expensive than running it again from scratch.

7.7 Appendix 7. Post Analysis Perspective Plots

Run Identifications

The Run Identification panel summarizes the number of identifications per run, displayed in a bar plot (Figure 75) and table format (Figure 76). On right-click you can change the basis of quantification between precursor, modified sequence, stripped sequence and protein group.

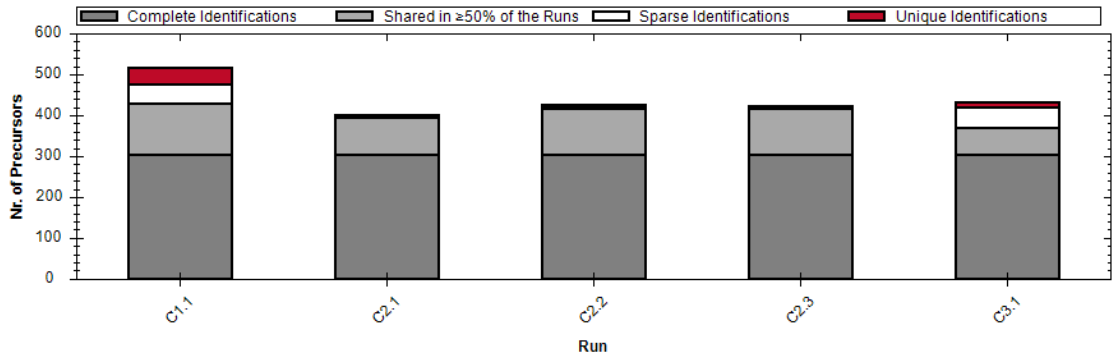


Figure 75. The upper-panel plot under the "Run Identifications" node shows the number of precursors per run and their completeness.

RunNr	FileName	Condition	Repli...	Precur...	ModifiedSeq...	Pepti...	ProteinG...	Proteins
E...	Contains:	Contains:	E...	Eq...	Equals:	E...	Equa...	Equals:
1	E_D180522_S509-SpecLip-DepPl-NoFrac_MHRM_...	S509-SpecLip-DepPl-No...	1	518	518	518	337	338
2	E_D180522_S509-SpecLip-Plasma-NoFrac_MHR...	S509-SpecLip-Plasma-N...	1	401	401	401	250	251
3	E_D180522_S509-SpecLip-Plasma-NoFrac_MHR...	S509-SpecLip-Plasma-N...	2	426	426	426	264	265
4	E_D180522_S509-SpecLip-Plasma-NoFrac_MHR...	S509-SpecLip-Plasma-N...	3	424	424	424	255	256
5	E_D180522_S509-SpecLip-ProteominerPl-NoFrac_...	S509-SpecLip-Proteomi...	1	434	434	434	277	278

Figure 76. The lower-panel table under the "Run Identifications" node shows the number of precursors per run and their completeness.

Data Completeness

In the data completeness plot, the "Cumulative Full Profiles" corresponds to identifications that were consistent across 1 to n runs, i.e. identified in all the runs currently looked at. On right-click you can change the basis of quantitation between precursor, modified sequence, stripped sequence and protein group.

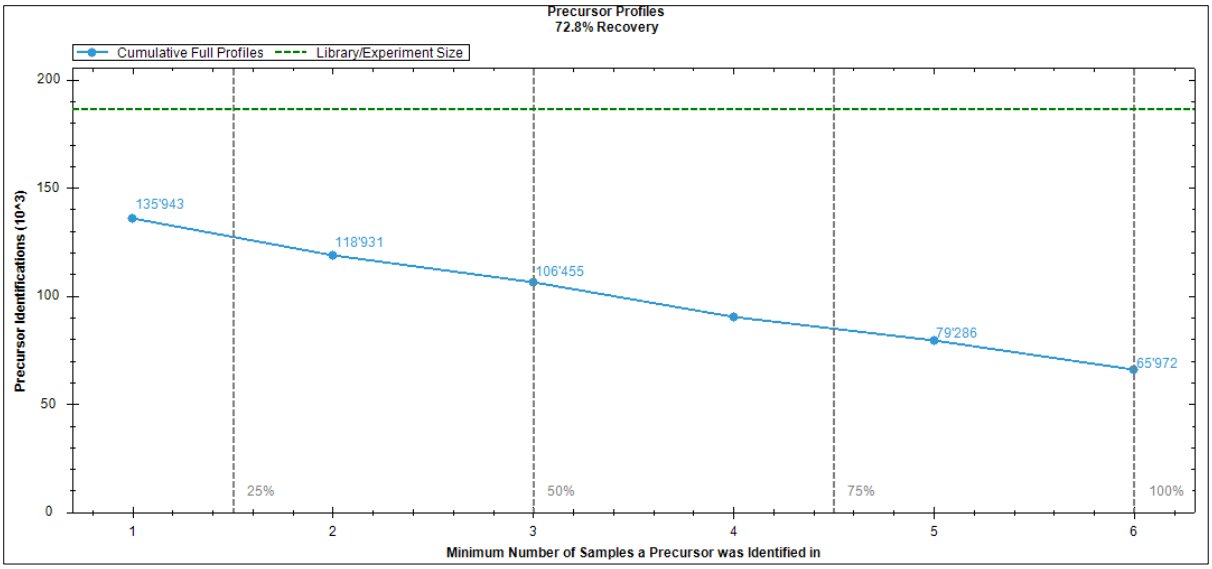


Figure 77. The upper plot under the "Data Completeness" node shows you number of cumulative full profiles. The blue line represent the decline of full profiles. The plot shows in an easy way how many precursors were identified in a given percentage of all runs. For example, in the first 3 runs (50% of the runs), 106.455 precursors where identified in all of those 3 runs. As with many other plots you can change the context to display the data completeness on precursor, peptide, protein-group or protein level.

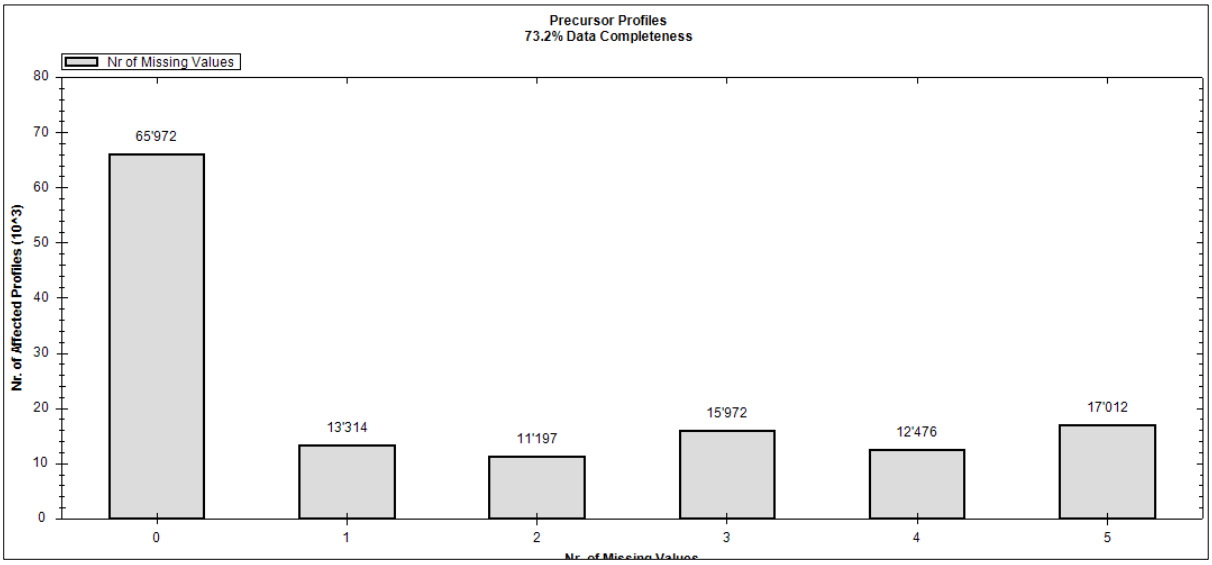


Figure 78. The lower plot under the "Data Completeness" node shows the histogram of missing values. In this plot, about 66.000 precursors did not have any missing values and where identified in every run (this is equivalent to the last data point in Figure 77) while about 13.300 precursors had one missing value and so forth. The last bar represents the number of single-hit precursors. These are the precursors that where only identified in one single run (hence they were only identified in 1 out of 6 runs in this case). As with many other plots you can change the context to show you the missing values on precursor, peptide, protein-group or protein level.

Ranked Protein Groups

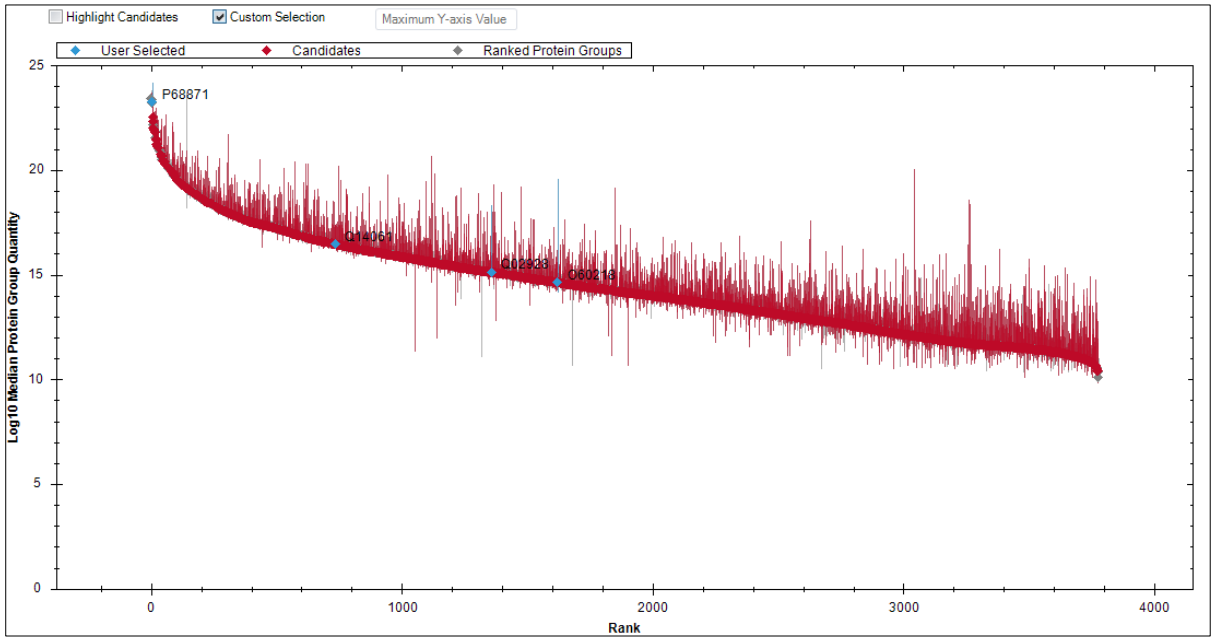


Figure 79. The Ranked Protein Groups plot shows all the identified protein groups ranked according to quantitative value. Quantitative variance indicators in red (interquartile range, IQR) , represent the distribution of protein abundances in different experimental runs. You can hover over each data point to find out which protein group correspond to each point. In addition, by selecting one of the boxes above the graph, you can choose to show labels of all candidates of the differential analysis or only of the selected ones.

Coefficients of Variations

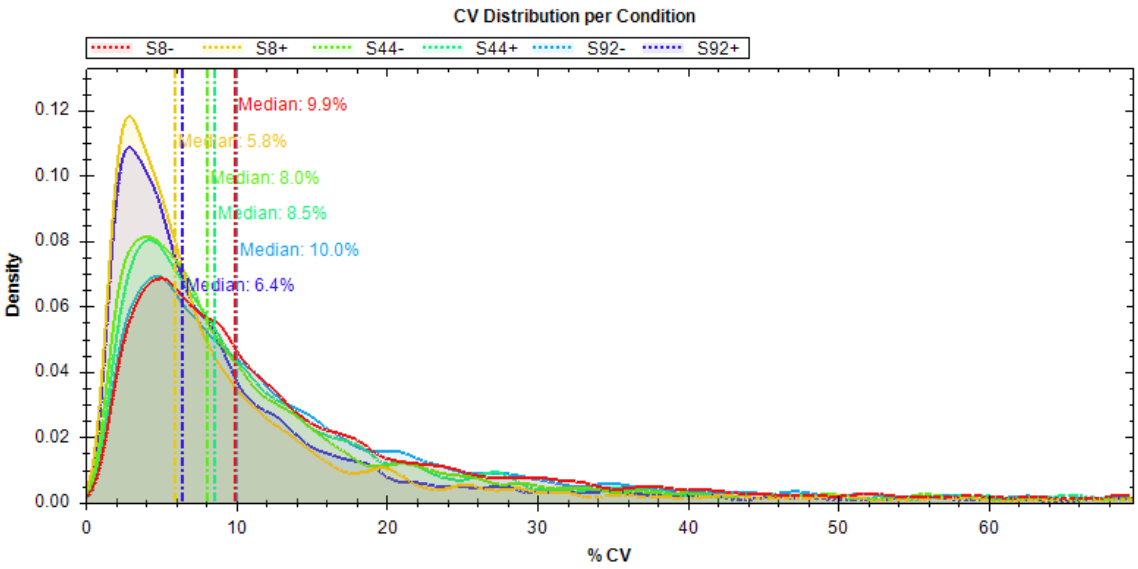


Figure 80. The upper plot under the "Coefficients of Variation" node shows the %CV distribution for all conditions in your experiment. You can also change the context by right-clicking on the plot to show you the CVs for precursor, peptide, protein-group and protein quantities or to show the CV distribution across all conditions.

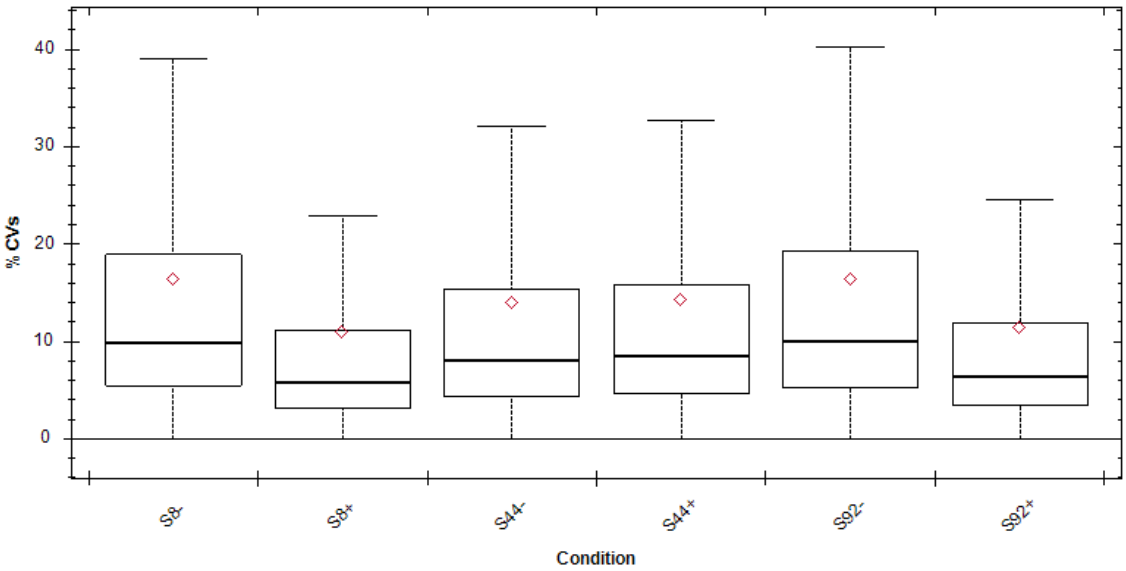


Figure 81. Similar to Figure 80, this plot shows the distribution of CVs for each condition. For experiments with large numbers of conditions, this view might be preferable. As with the previous figures, you can change the context of the plot by right-clicking on it to select either precursor, peptide, protein-group or protein scope. Data could be displayed as box plot or as a violin plot.

CVs Below X

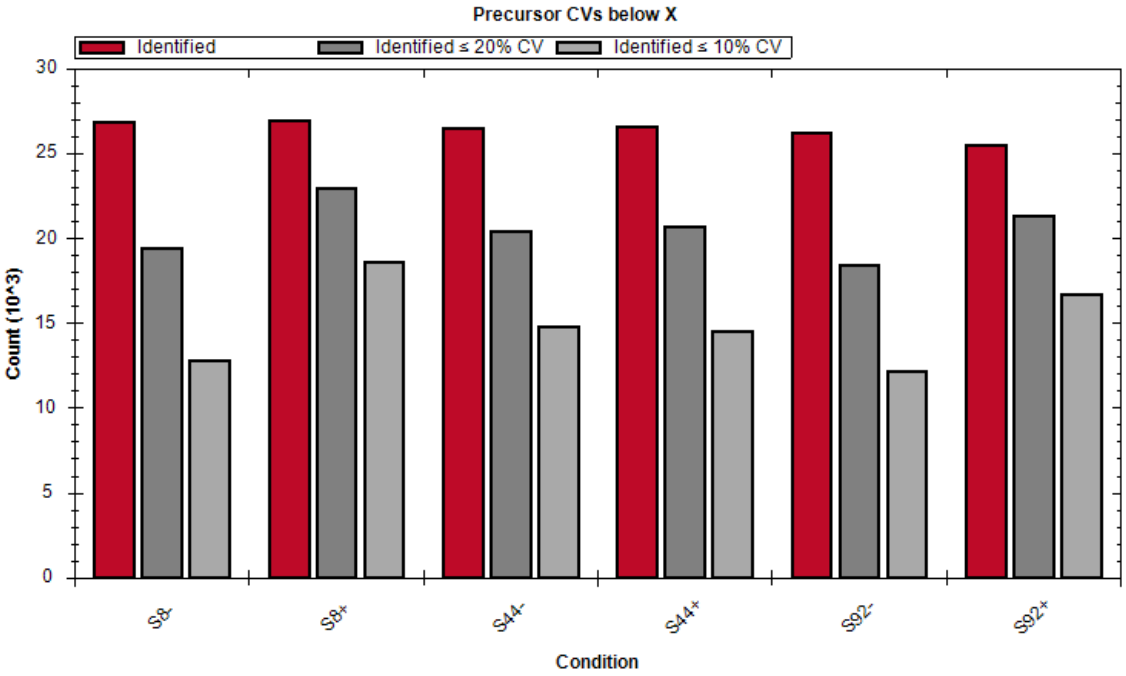


Figure 82. The CVs below X plot shows the number of precursors that were below either 20% or 10% CVs. The red bar shows the number of all identifications for each condition regardless of quantitative precision. The bar for ≤ 20% CVs also includes all counts from ≤ 10% CVs. As with the previous figures, you can change the context of the plot by right-clicking on it to select either precursor, peptide, protein-group or protein scope.

Normalization

In the normalization you can see boxplots of responses for the individual runs before and after normalization.

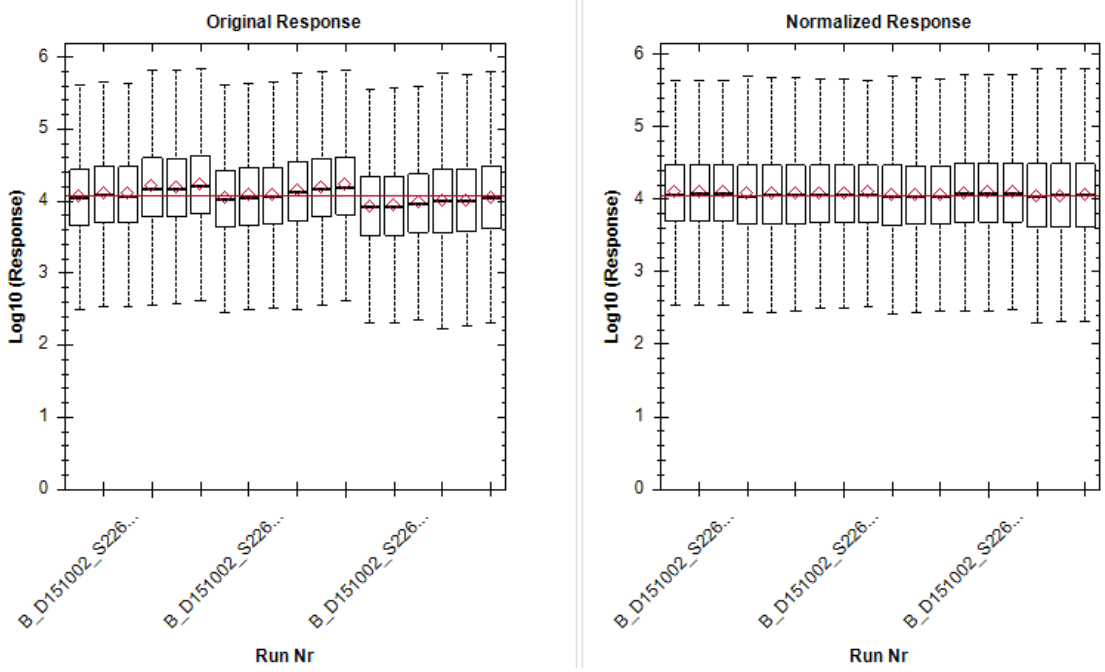


Figure 83. The "Normalization" node shows details about the normalization status of you experiment. The left side shows boxplots of precursor quantities before normalization for each run. The right side shows boxplots of the same precursor quantities after normalization. After right clicking on the graph, you can display data as a violin plot.

Coefficients of Variation

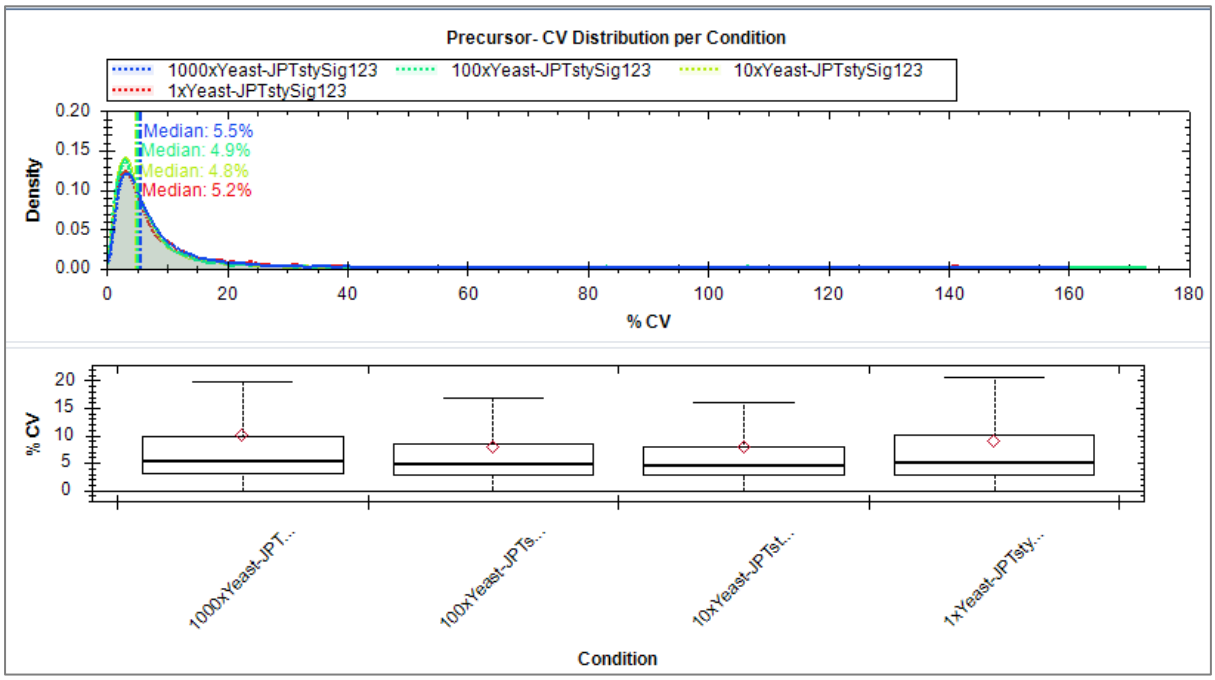


Figure 84. Coefficients of Variation plot shows distribution of coefficients of variation in each of the experimental conditions. By right clicking at the plot, the CV base could be changed from precursor to peptide, protein or protein group. Also, you can select to show data per sample or per whole experiment.

Datapoints Per Peak

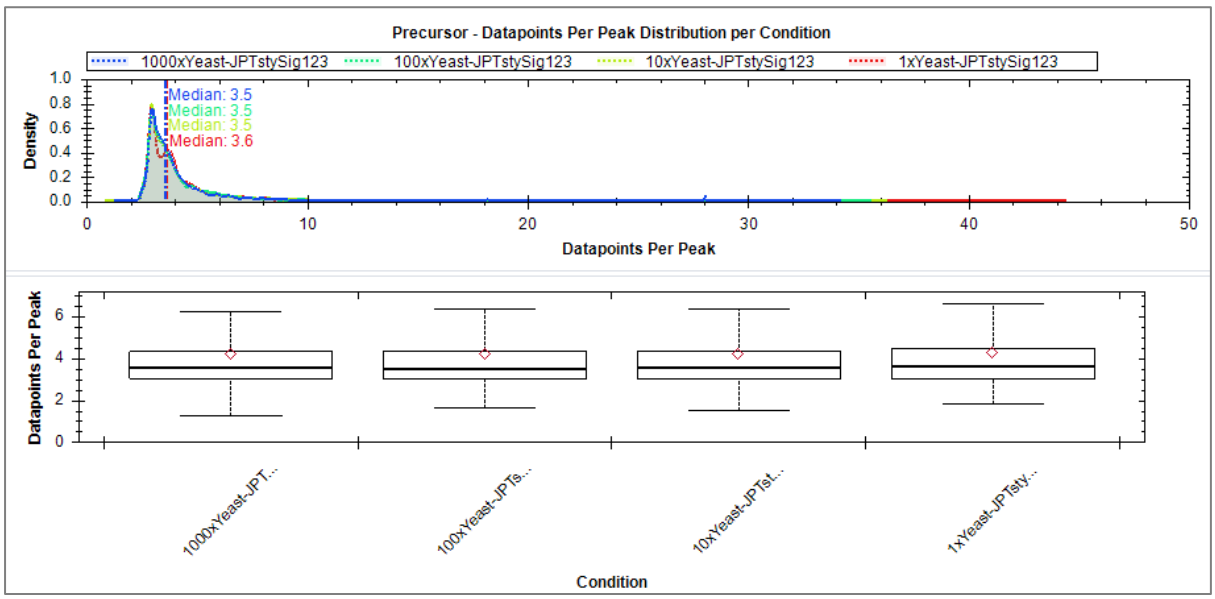


Figure 85. Datapoints per Peak plot shows the data points per peak in each condition as well as their distribution with the median. By right clicking, the grouping of the data could be changed to grouping per sapmple, per condition or across whole experiment.

Binned Identification

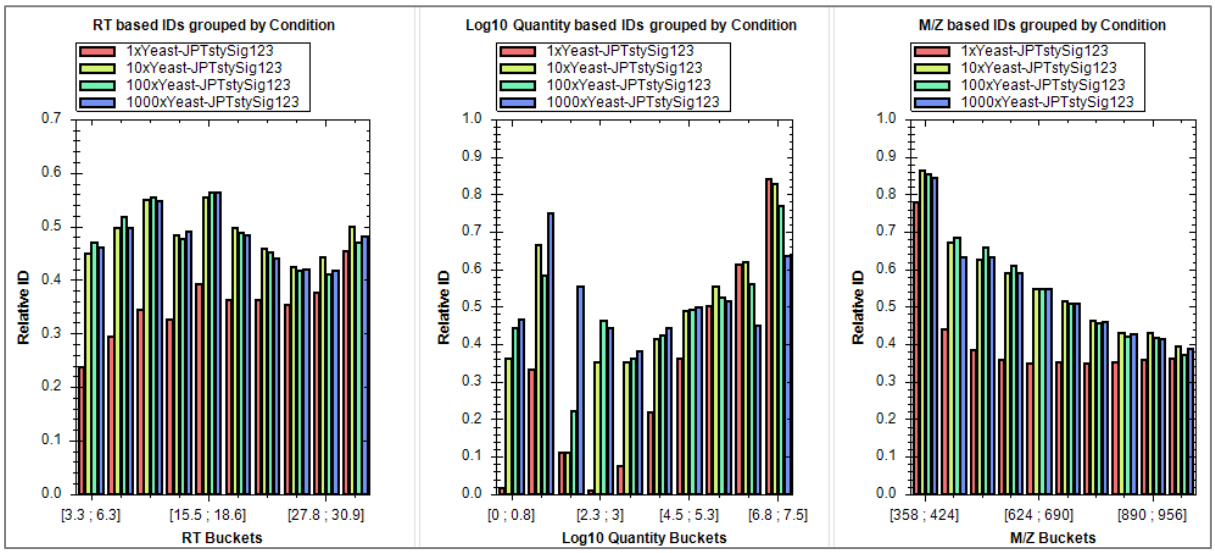


Figure 86. Binned Identification plot shows relative ID plotted for Retention Time buckets, Log10 Quantity buckets and m/z buckets. Data is grouped by the condidtion which, upon right clicking, could be changed to run or fraction.

LFQBench

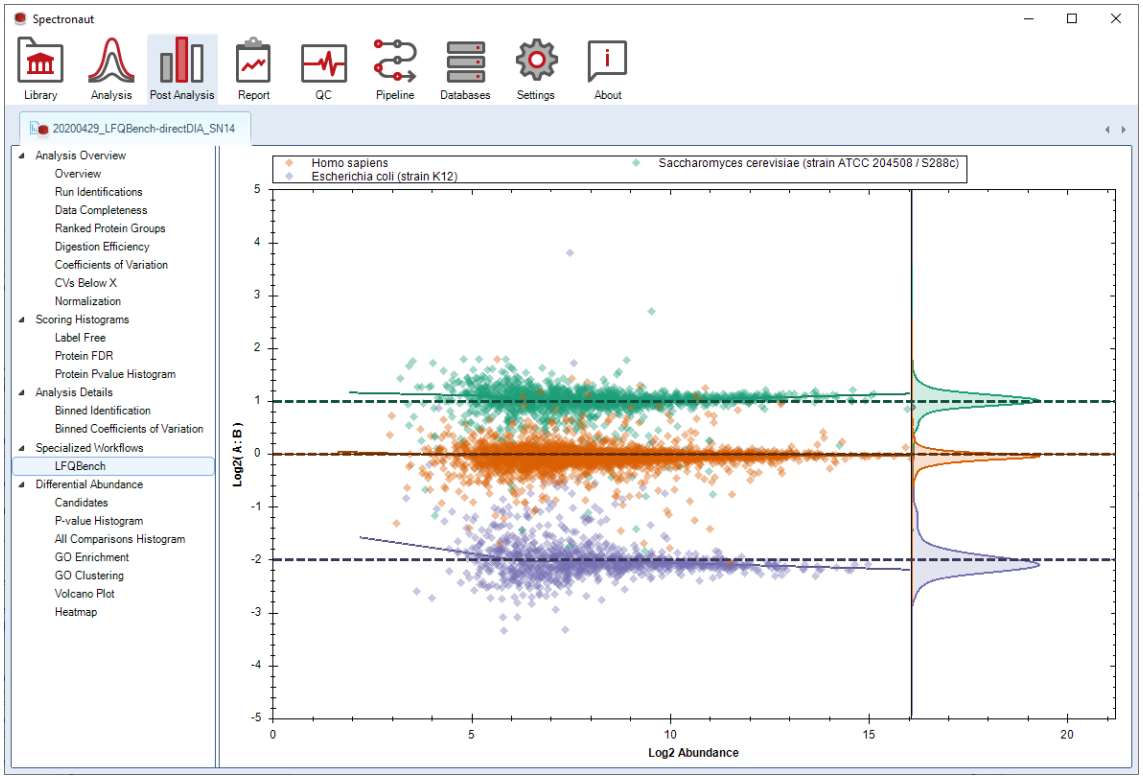


Figure 87. The LFQBench plot for mixed proteomes of three organisms. The Y-axis (log2Ratio) shows the ratios of the mixed proteomes based on experimental data. The expected ratios (dashed lines) can be set by

right-clicking on the plot. The distributions on the right side of the plot show how the experimental diverge from the expected ratios. The X-axis represents protein abundance (log2-transformed).

Heatmap

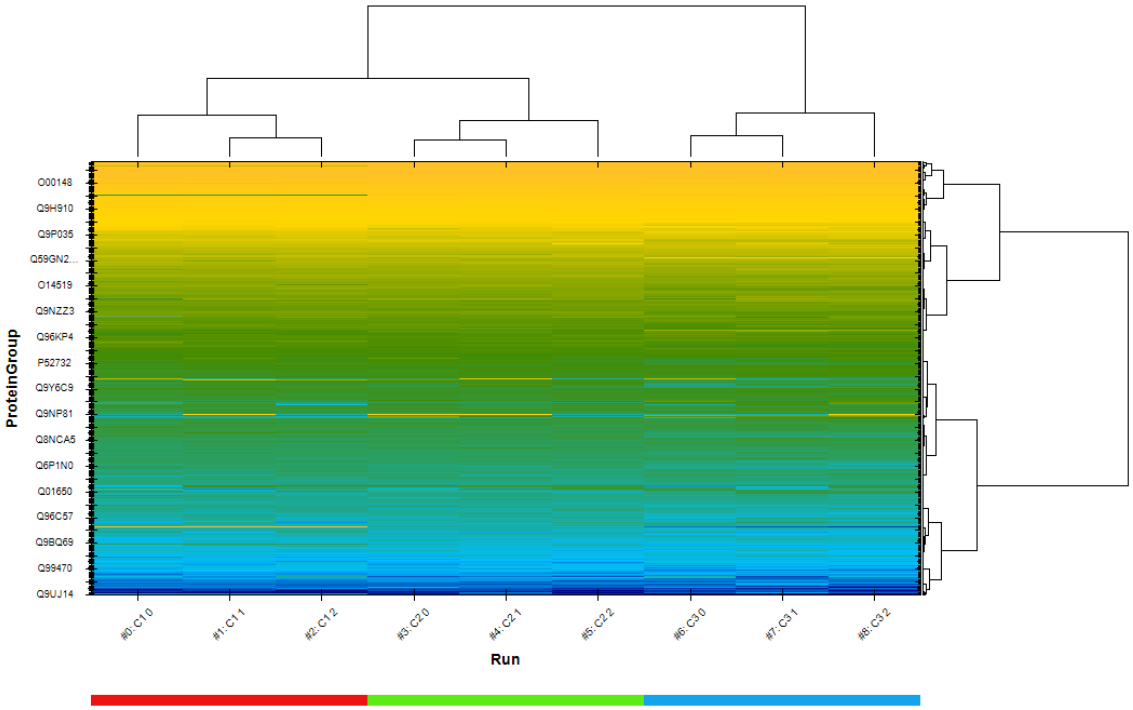


Figure 88. Heatmap showing the clustering of 9 runs from 3 conditions. Runs within the same condition cluster nicely as illustrated by the condition-based color code in the bottom of the heatmap and the x-axis dendrogram.

Volcano Plot

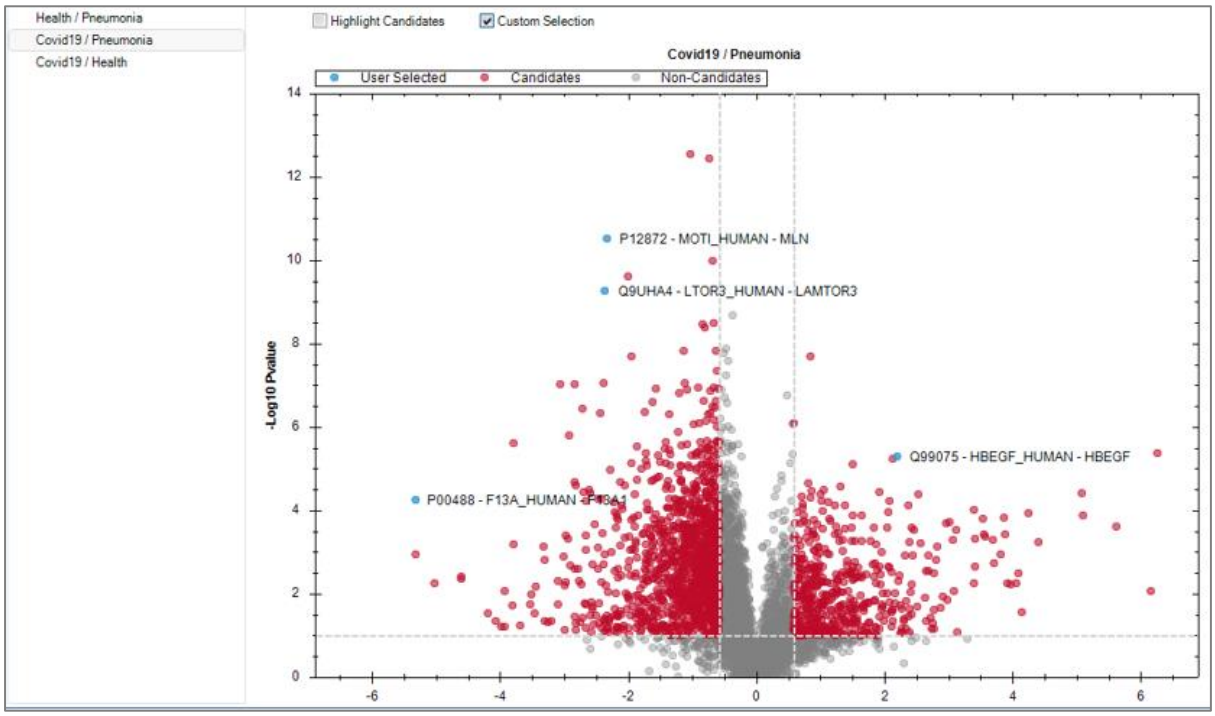


Figure 89. Volcano plot showing the potential candidates of an experiment containing 3 conditions to each 3 replicates. By default, the filters are set to ≥ 1.5 absolute fold change and ≤ 0.05 q-value. Selection of Highlight Candidates box above the graph will show labels of all candidates. Custom Selection can be chosen to highlight only selected ones.

Sample Correlation Plot

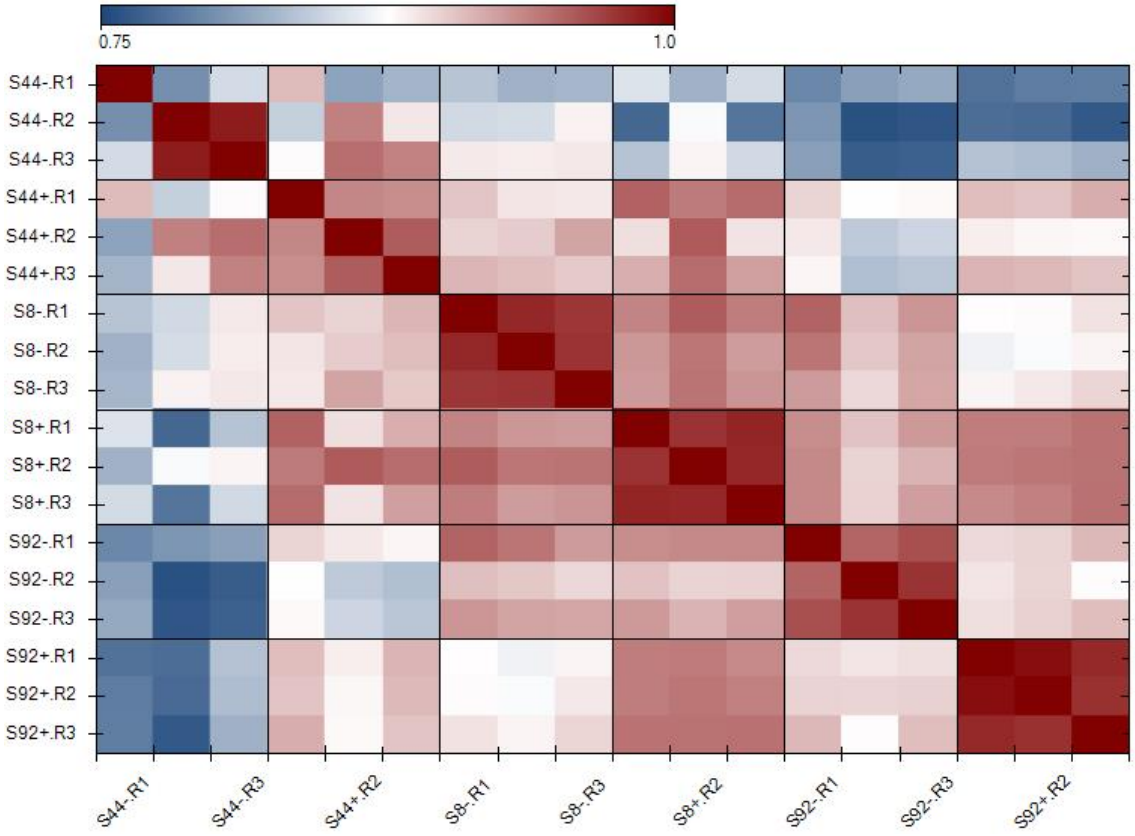


Figure 90. The sample correlation matrix shows correlation of precursor quantities between all samples. The rows and columns of the matrix are ordered by condition and replicate annotation. A high correlation between technical replicates of the same sample is to be expected while low correlation between different samples might indicate biological variance. By default, the coloring range is set from 0.75 to 1.0 but can be changed via right-click option.

PTM Analysis Plots

PTM Analysis plots in the Post Analysis Perspective are available when the PTM workflow is selected for the analysis of a given experiment.

Principal Component Analysis

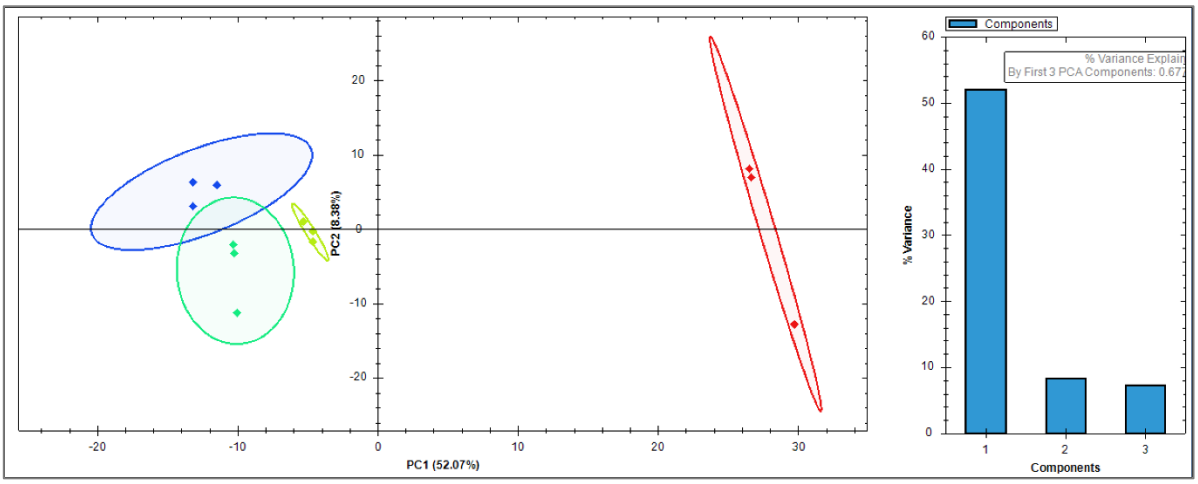


Figure 91. Principal Component Analysis plot shows clustering of the samples based on their modification sites profiles. The components bar plot on the right side shows how the first three components explain the variance.

Volcano Plot

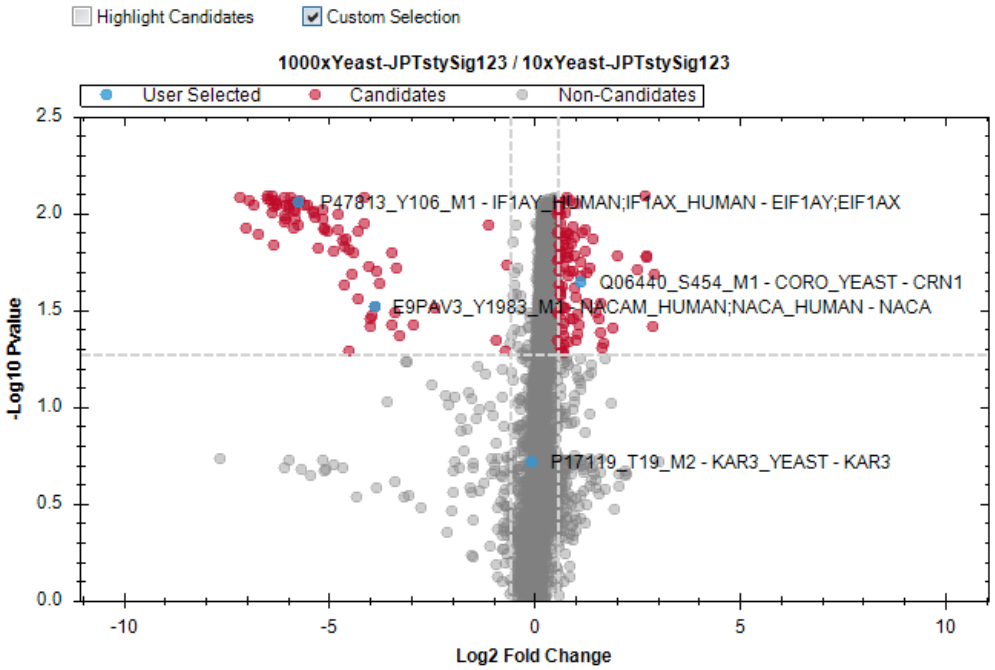


Figure 92. Volcano Plot of the PTM analysis shows -Log10 P value plotted against Log 2 Fold change of differential abundance on the modified site level. By default, the filters are set to ≥ 1.5 absolute fold change and ≤ 0.05 q-value. Selection of Highlight Candidates box above the graph will show labels of all modified sites candidates. Custom Selection can be chosen to highlight only user selected ones.

PTM vs Protein Fold Changes

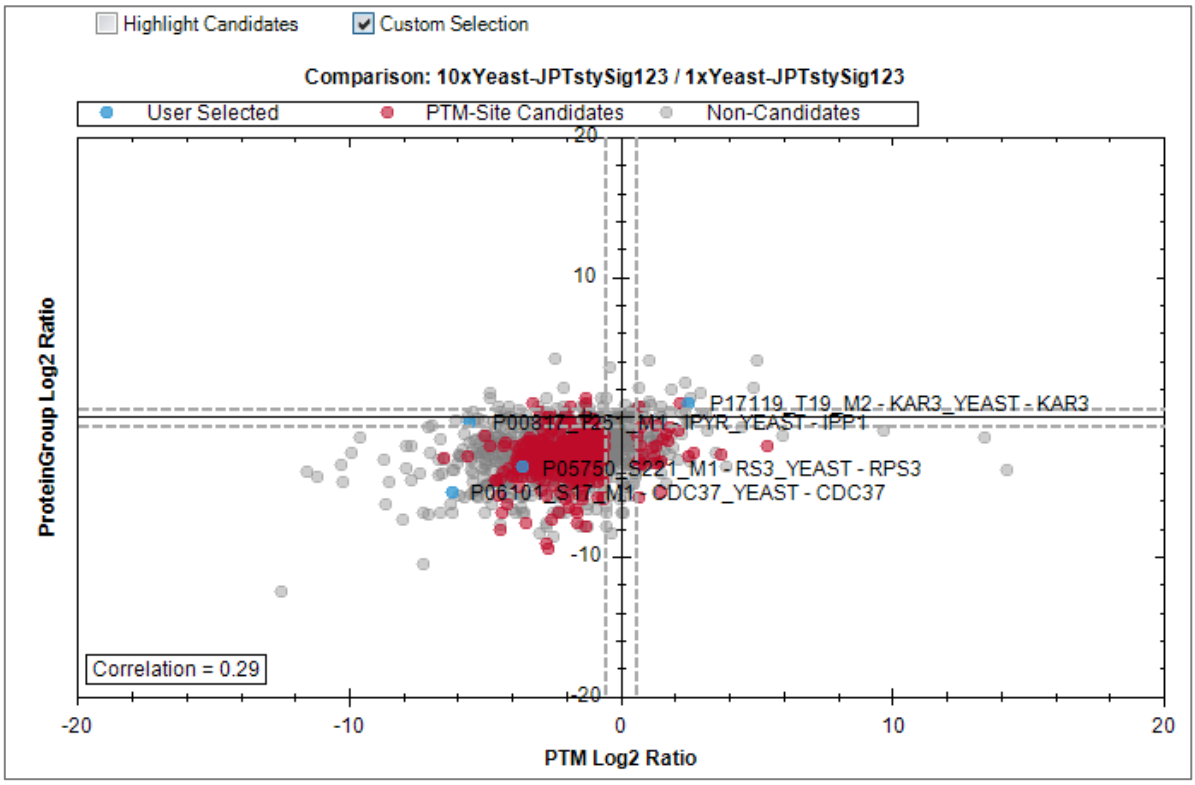


Figure 93. PTM vs Prtoein Fold Changes plot shows log2 ratio of the protein groups ploted against Log2 ratios of the PTM sites. Selection of Highlight Candidates box above the graph will show labels of all PTM differential analysis candidates. Custom Selection will enable highlighting only candidates of intrest.

Modification Enrichment

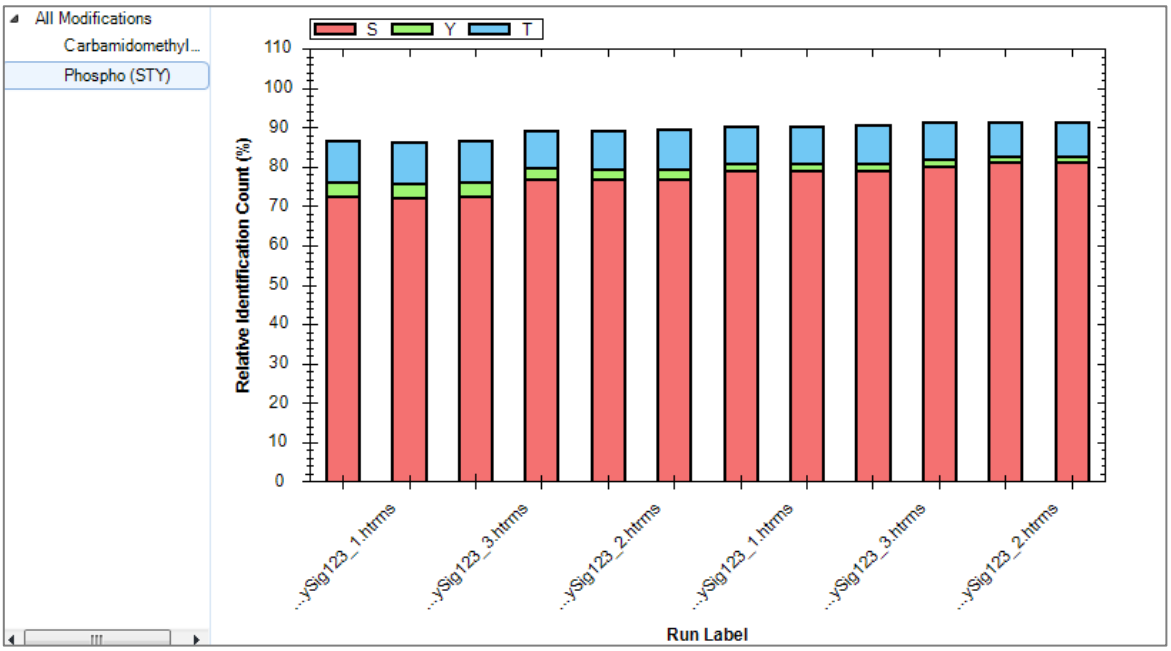


Figure 94. Example of the Modification Enrichement plot showing the percentage of the identified precursors that are carrying a phosphorylation at particular amino acid in each of the runs. The plot can display relative identification count (as on the graph above) or the quantity.

7.8 Appendix 8. Most Relevant Report Headers

Many headers have a text hover tool directly in the software. If you don't find the information you are looking for, do not hesitate to contact us via our Help Center.

Protein Group (PG) headers

Headers related to Protein Group (PG) as defined in the settings. Most headers related to PGs are self-explanatory. Here are the most relevant and some which are not too obvious

PG.ProteinGroups	One or several protein groups separated with a "|". Protein ids within protein groups are separated with a ";". The protein groups can either originate from the Spectronaut IDPicker protein grouping or from the search engine used to generate the spectral library.
PG.Quantity	The quantitative value as defined in the settings.
PG.Qvalue	The q-value (FDR) for that PG. The q-values for protein groups are experiment-wise.
PG.RunEvidenceCount	Number of hits (precursors) found for that PG in that run.
PG.IsSingleHit	True of False. It tells you whether the PG was found with only one hit, as defined in the settings.
PG.ProteinAccessions	The protein accessions in the same order as the protein ids in "PG.ProteinGroups". The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported
PG.Genes	The genes in the same order as the protein ids in "PG.ProteinGroups". The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported
PG.ProteinDescriptions	The protein descriptions in the same order as the protein ids in "PG.ProteinGroups". The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported

PG.UniProtIds	The UniProt ids in the same order as the protein ids in "PG.ProteinGroups". The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported
PG.ProteinNames	The protein names in the same order as the protein ids in "PG.ProteinGroups". The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported
PG.FastaHeaders	The FASTA headers in the same order as the protein ids in "PG.ProteinGroups". The value corresponds to what was defined by the parsing rule. If all the values are identical a unique value is reported

Peptide headers

Headers related to Peptides (PEP) as defined in the settings. Many headers related to Peptides are self-explanatory. Here are the most relevant and some which are not too obvious.

PEP.GroupingKey and PEP.GroupingKeyType	Tells you which element is considered as peptide (defined in the settings). Default is stripped sequence.
PEP.IsProteinGroupSpecific	True or False. Tells you whether the peptide only belongs to one Protein Group.
PEP.RunEvidenceCount	Number of hits (precursors) found for that peptide in that run.
PEP.Quantity	The quantitative value for that peptide as defined in the settings.
PEP.UsedForProteinGroupQuantity	True or False. Tells you whether this peptide was used to calculate PG quantities, as defined in the settings.

Elution Group (EG) headers

Headers related to Elution Groups (EG). Many headers related to EG are self-explanatory. Here are the most relevant and some which are not too obvious.

EG.PrecursorId	Unique Id for the precursor: [modified sequence] plus [charge]
EG.Library	The library this assay is from
EG.Workflow	The workflow this EG was used in. It can be LABEL_FREE, SPIKE_IN or LABEL
EG.IsUserPeak	True or False. Tells whether the EG was manually integrated
EG.Identified	True or False. The EG has to pass the precursor and protein q-value cutoff to be considered identified
EG.Qvalue	The q-value (FDR) of the EG.
EG.PTMAssayCandidateScore	Score of the Localization candidate that corresponds to the assay
EG.PTMAssayProbability	The probability that the localization of the assay corresponds to what was observed in the data
EG.PTMLocalizationProbabilities	All localization information of the peptide in sequence annotation
EG.PTMPositions [Mod-Name]	The sequence positions for all amino acids that could carry this modification
EG.PTMProbabilities [Mod-Name]	The probabilities for the individual positions
EG.PTMSites [Mod-Name]	The amino acids that correspond to the modifiable positions of this sequence
EG.ProteinPTMLocations	Modified locations within the parent protein sequence(s)
EG.TotalQuantity	The quantitative value for that EG as defined in the settings.
EG.Label	A label for a peptide precursor compiled of the protein id, stripped sequence and the precursor ion charge. The label is not necessarily unique and therefore not intended for downstream peptide profiling across runs

EG.HeightCV	In a multi-run experiment, this value reports the CV of the peptide based on the summed apex height of the detected peak
EG.AreaCV	In a multi-run experiment, this value reports the CV of the peptide based on the summed area under the curve of the detected peak
EG.iRTEmpirical	The iRT (Escher et al. 2012) as determined in this specific analytical run
EG.MeanApexRT	The average retention time of the peak apexes across all fragment ions of this peptide
EG.MeanTailingFactor	The average tailing factor of the elution group across all the fragment ions determined at the FWHM
EG.DeltaiRT	The difference between the predicted and empirical iRT. A measure of the reproducibility of chromatography
EG.IsVerified	Reports manual accepting or rejecting of peaks assigned in the Analysis Perspective
EG.Cscore	The Spectronaut identification score, which is based on an advanced mProphet (Reiter et al. 2011) scoring. A high score indicates high quality identifications
EG.IsUserPeak	Specifies whether the peak was integrated automatically or manually in the Analysis Perspective
EG.AllProteinAccessions	All protein accessions this peptide points to. This field is only reported when the Spectronaut protein inference was used. It represents the input to the IDPicker protein grouping algorithm

Fragment Group (FG) headers

Headers related to Fragment Group (FG). FG is only relevant in labeled and spike-in workflows. Two FGs belong to one EG. The FG id corresponds to the EG id plus the isotopic labelling.

Many headers related to FG are self-explanatory. Here are the most relevant and some which are not too obvious.

FG.Id	A unique ID of the peptide precursor. Corresponds to the EG.Id if the experiment is label free
FG.Label	A label of the fragment ion group. The label is not necessarily unique and therefore not intended for structuring data
FG.Charge	The charge state of the peptide precursor
FG.PrecMz	The peptide precursor m/z
FG.PrecWindowNumber	The precursor window number in which the precursor was measured
FG.TotalPeakArea	The summed-up peak area of all fragment ions for the corresponding peptide precursor
FG.TotalPeakHeight	The summed-up apex peak height of all fragment ions for the corresponding peptide precursor
FG.Quantity	The quantitative value for that FG as defined in the settings.

Fragment (F) headers

Headers related to Fragment ions. If you choose fragment ion level information in your report, you will have one row per fragment ion. This can make the report considerably large.

Most headers related to Fragments are self-explanatory. Here are the most relevant and some which are not too obvious.

F.InterferenceScore	Fragment ion interference score as determined by Spectronaut
F.PossibleInterference	True or False. Tells you whether an ion is a probable interference.

F.ExcludeFromQuantification	True or False. An ion can be excluded via the library or of it is considered an interference.
F.NormalizePeakArea	The quantitative value calculated as the area under the curve

PTM site headers

PTM site report is available whenever the experiment was analyzed with PTM workflow. The most relevant PTM site headers are listed and described below.

PTM.CollapseKey	The key used to group and condense parent peptides together into the PTM site object.
PTM.FlankingRegion	The flanking region of amino acids around the site location modified with given PTM
PTM.Group	All parent peptide sequences that were condensed in the PTM site object
PTM.ModificationTitle	The title of the modification that is presented by this PTM site object
PTM.Multiplicity	The maximum multiplicity observed for this PTM site. The multiplicity is defined by the number of the modifications of the same type being identified at any of the parent peptides containing this modification
PTM.NrOfCollapsedPeptides	The number of peptides that were condensed into this PTM site object.
PTM.ProteinId	The parent protein ID for this PTM site.
PTM.Quantity	The condensed quantity of all parent peptides that cover this PTM site.
PTM.SiteAA	The amino acid that PTM site is modifying.
PTM.SiteLocation	The amino acid sequence position of this PTM site in the parent protein sequence.
PTM.SiteProbability	The highest observed site probability corresponding to this PTM site in this run/ sample.

7.9 All XIC database export

Spectronaut 18.5 introduced the option to export all XICs, generated during the main quantitative DIA search, in an SQLite database format. The full XIC db export can be triggered

1. Via a right-click option on the experiment tab in the Review perspective (Export all XIC)
2. Via a global export rule (Settings -> Global -> Reporting -> Automatic XIC Storage
3. Via the analysis pipeline sub-settings when run from command line (Analysis Settings -> Pipeline Mode -> Export All XICs)

Database Schema

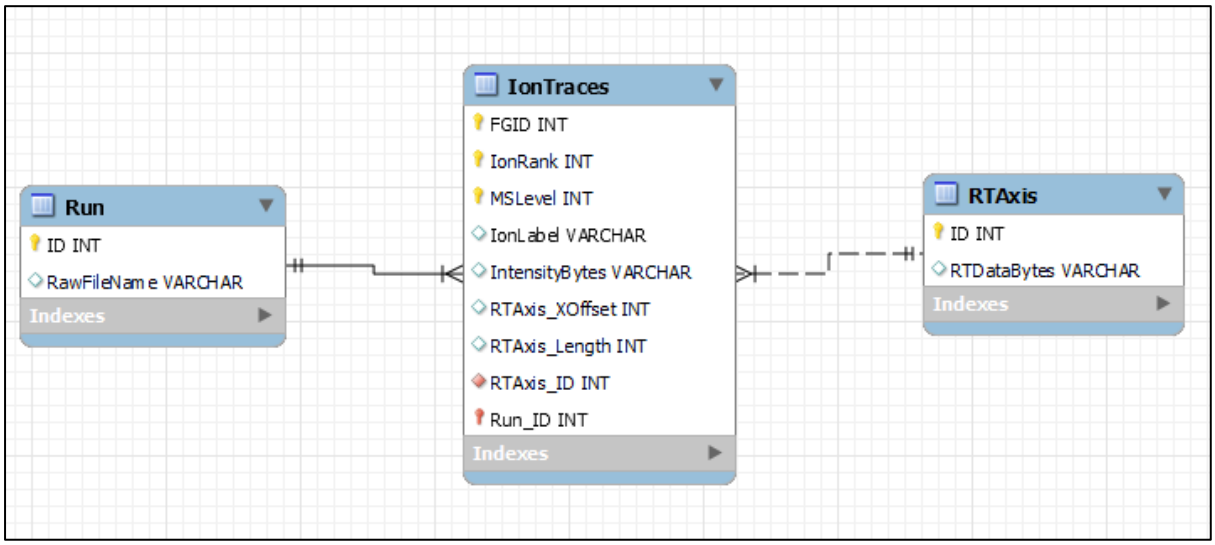


Figure 95 XIC export database schema. The XIC db can be exported on a per-experiment basis. The exported database consists of 3 SQL tables.

IonTraces Table

The IonTraces table is the main data-table of the XIC database. It consists of the unique link between the FG.XICDBID field from the Spectronaut report to the FGID database key and all relevant meta information to visualize simple plots.

- FGID: The database identifier that uniquely identifies the parent precursor from a Spectronaut report (FG.XICDBID).
- IonRank: The SN ranking of the ion. This is used as additional primary key as the rank of fragments is always unique (unlike the fragment label).
- MSLevel: In combination with the other two primary keys, the addition of the MSLevel is now sufficient to uniquely identify any ion from a Spectronaut output.

- IntenstiyBytes: This field contains the XIC intensities as 32bit (4byte) floating point array in base64 string encoding.
- RTAxis_XOffset**:** Specifies the offset index to find the corresponding RT values for these intensities from RTAxis table.
- RTAxis_Length**:** Specifies the number values to read from the RTAxis array corresponding to the length of the intensity float[] array.
- RTAxis_ID: Identifies the correct RT axis to be used from the RTAxis table. The table consists of 1 RTAxis array per scan-group.
- Run_ID**:** Links the current row to the parent run from the Run table.

RTAxis Table

This table contains the corresponding RT axis for the stored XICs. The RT information is fully redundant between fragments of the same precursor, and highly redundant between fragments within the same scan-group (SWATH). This is why the RT dimension for the entry group is stored as one entry corresponding to all scans that belong to this group (SWATH).

- ID: Unique identifier for a given RTAxis space.
- RTDataBytes: This field contains the full RT dimension for the corresponding scan-group (SWATH). Like the IntensityBytes, the data is stored as a base64 encoded 32bit (4byte) floating point array and can be converted back the same way. In order to retrieve the correct RT dimension range, the RTAxis_XOffset and RTAxis_Length fields from the Iontraces table are used.

Run Table

This table holds the information about the source raw-file for each stored iontrace. The raw files are stored based on their order in the original experiment.

- **ID**: Unique identifier of a raw-file within the exported experiment. It corresponds to the run-index at time of exporting.
- **RawFileName**: The full file name of the corresponding raw file including the file ending.

Appendix 9. Read from XIC export DB (example code in R)

```
# Install and initialize relevant packages
install.packages("RSQLite")
install.packages("base64enc")

library(RSQLite)
library(base64enc)

fgid <- 220083 #Precursor DB ID of interest
runid <- 0
mslevel <- 2 #Select all MS2 ions of the selected precursor


db <- dbConnect(SQLite(), dbname = [PATH-TO-XIC.DB])
queryTemplate <- "SELECT * FROM IonTraces INNER JOIN RTAxis ON
Iontraces.RTAxis_ID=RTAxis.ID WHERE FGID='%s' AND Run_ID='%s' AND MSLevel='%s'"

query <- sprintf(queryTemplate, fgid, runid, mslevel)
result <- dbGetQuery(db, query)

#Read XIC intensity and RT data from results
xicData <- result[["IntensityDataBytes"]]
rtData <- result[["RTDataBytes"]]

#Read sub range information for RT dimension
offsets <- result[["RTAxis_XOffset"]]
lengths <- result[["RTAxis_Length"]]

# Split the base64 data into individual rows
base64_XICData <- unlist(strsplit(xicData, ","))
base64_RTData <- unlist(strsplit(rtData, ","))

# Initialize an empty list to store the float arrays
intensities <- list()
rts <- list()

# Read and decode the Intensity values
for (i in 1:length(base64_XICData)) {
  base64_string <- base64_XICData[i]
  decoded_data <- base64decode(base64_string)
  num_floats <- length(decoded_data) / 4

  #Convert binary data back to 4byte floating point array
  float_array <- readBin(decoded_data, what = "numeric", n = num_floats, size =
4, endian = "little")
  intensities <- c(intensities, list(float_array))
}

#Continue on next page
```

```
# Read and decode the RT dimension
for (i in 1:length(base64_RTData)) {
  base64_string <- base64_RTData[i]
  decoded_data <- base64decode(base64_string)
  num_floats <- length(decoded_data) / 4

  #Convert binary data back to 4byte floating point array
  float_array <- readBin(decoded_data, what = "numeric", n = num_floats, size =
4, endian = "little")

  #For the RT dimension, we must extract the relevant sub array
  startoffset <- offsets[i]+1 #R indexing starts at 1
  endOffset <- (startoffset + lengths[i])-1
  sub_array <- float_array[startoffset:endOffset]
  rts <- c(rts, list(sub_array))
}

# Create a single plot for all float arrays
plot_title <- "XIC"

# Plot all float arrays on the same plot
plot(rts[[i]], intensities[[1]], type = "l", main = plot_title, xlab = "Retention
Time", ylab = "Intensities", col = 1)

# Add lines for the rest of the float arrays
for (i in 2:length(intensities)) {
  lines(rts[[i]], intensities[[i]], col = i)
}


dbDisconnect(db)
```