SpectroMine™

powered by Pulsar

User Manual

1 General Information

1.1 Scope of SpectroMine™ Software

SpectroMine™ is a commercial software package aimed at processing and analyzing a large variety of mass spectrometric proteomic data in different ways:

1. **Search**: SpectroMine comes with Pulsar – integrated Biognosys' proprietary search engine. Pulsar can search data-dependent acquisition (DDA), data-independent acquisition (DIA) and parallel reaction monitoring (PRM) data with MS1 information. Both centroid and profiled data can be processed. Pulsar is designed to be fast and scale efficiently as the number of runs increase. 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.

   The specific vendors and acquisition modes supported by Pulsar are:

   - Thermo Scientific™ (DDA, DIA and PRM, FAIMS DDA and DIA; also data acquired with EthcD fragmentation)
   - Sciex (DDA and DIA/SWATH™)
   - Bruker (DDA, DIA, PASEF, dia-PASEF)
   - Waters (DDA, HDMS^E)

   The algorithm followed by Pulsar to search DIA data is based on DIA-Umpire (Tsou *et al.* 2015).

   The search results can be explored and saved. You can also generate a library from these searches, as well as save the Search Archive to explore your results later, or to use it for library generation (Box 1).

2. **Quantify**: SpectroMine can analyze DDA data from isobaric label quantification (ILQ) and label-free quantification (LFQ) experiments. SpectroMine can quantitatively profile several 1000s of proteins in one experiment. The number of samples in an ILQ experiment is limited by the number of reporter ions available for a particular isobaric tag; however, SpectroMine is developed to support block experimental designs, which can significantly increase the multiplexity of the experiment (Box 2). Finally,

SpectroMine will statistically assess differential abundance and will report candidates that significantly change between conditions.

SpectroMine is optimized to analyze data from all ion trap resolutions. MS1, MS2, and SPS-MS3 quantification are supported. In addition, neutral loss-based fragmentation label methods (such as EASI-tag) are supported. Fractionation, including gas-phase fractionation is supported for ILQ experiments.

1.2 SpectroMine™ 3.0 Key Release Features

- ❖ Pulsar Database Search Engine
  - Up to 70% faster for immunopeptidomics and many modification searches
  - Up to 25% faster for standard searches
  - Improved performance for large protein databases (> 2 GB) with Pulsar
- ❖ Accurate Quantification
  - Improved default settings for quantification
  - Improved impurity correction for TMTpro
  - Improved turboTMT support
  - New MaxLFQ based protein quantification
  - New TMTpro 18-plex support
- ❖ Comprehensive PTM Analysis
  - Site collapse calculations
  - Site regulation analysis
  - New PTM site report
- ❖ Improved Immunopeptidomics support
  - Up to 50% faster processing of immunopeptidomics data
  - Motif plot visualization for identified n-mers
- ❖ Improved Deep-learning Augmentation
  - Prediction for ion mobility (1/K0) during library generation
  - Improved fragment prediction for unspecific peptides
- ❖ New Post Analysis Perspective Features
  - Added PCA analysis
  - Added custom selection for volcano plots
  - New Run based Overview node (Identifications, Peptide Length Distributions, Peptide Motif Plot, and ILQ NA Ratio plot per channel)
  - New Summarized Fractions node (Quantifications Overview)

- New PTM Analysis node (PTM Site-based differential analysis plots)

❖ Improved User Experience and Visualization

- Improved Protein Coverage plot
  - PTM support
  - Peptide tooltip and highlighting
- Detachable perspective and plots for side-by-side visualization
- Improved plot and tree filter selection
- Quick-Action bar for analysis perspective
- Improved warning messages with actionable links
- Improved UI for analysis log

❖ Improved Report Perspective

- New Summarized Fraction report when analysing a fractionated experiment
- New PTM Site level report when performing PTM localization

1.3 Computer System Requirements

SpectroMine™ is only available for Windows operating systems. Command line operation is also supported (see section 3.10). The minimum and recommended system specifications are:

Minimum

Operating System:	Windows 10, x64
CPU:	Intel® Core™ CPU, 2.7 GHz (quad-core) or similar
Hard drive:	200 GB free space
Memory:	16 GB
Software:	.NET Framework 4.7

Recommended

Operating System:	Windows 10 or higher, x64
CPU:	Intel Core i7 4770, 3.4 GHz (octa-core) or similar, Intel or AMD CPU with 4 or more cores
Hard drive:	2 TB of hard drive space or more (2x data set size), solid state drive (SSD)

Memory: 64 GB or more

Software: .NET Framework 4.7 or higher

Box 1. Search Archives

Search Archive is a new concept introduced by Biognosys.

Before Search Archives, previously 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 cost a great amount of time and computational resources which had to be reinvested.

With Search Archives, every time you search with Pulsar, the complete result relevant to that Pulsar search is saved as a Search Archive and will appear in the Search Archive page of the Archive Perspective. Search Archives contain the information from a search before applying any FDR filter. This allows the combination of several Search Archives, or Search Archives with runs files, to generate libraries with a proper, library-wide control of the FDR.

There are two types of Search Archives:

1. Search Archives from run files (complete Search Archives). These archives contain all the information related to the Pulsar search of each particular run file.
2. Search Archives from other archives or from combinations of archives (meta-Search Archives). These archives contain the metainformation of:
   - A re-utilized complete archive
   - A combination of archives

When a library is generated combining archives and run files, a complete Search Archive will be stored for each run file and a meta-Search Archive will be done containing the information of the combined analysis.

When meta-Search Archives are used, the process will be directed to the corresponding complete Search Archives to retrieve the relevant information.

1.4 Post installation recommendations

Performance improvements after the installation:

1. **Directories:** SpectroMine™ 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 directories to a local destination with enough free memory. To do that, go to the Settings Perspective → Global → Directories

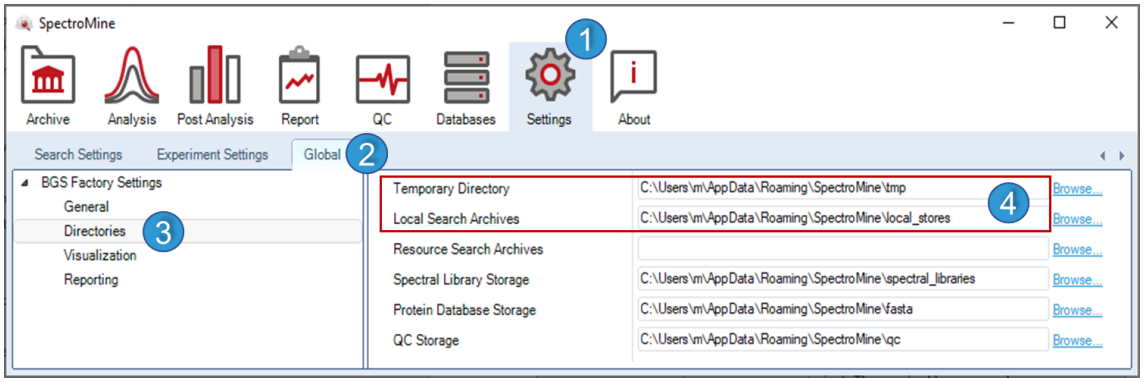

Figure 1. Change the default location for temporary files and Search Archives.

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

3. **CPUs:** SpectroMine is designed to perform highly resource intensive tasks. For resource management purposes, you can set a maximum number of CPUs SpectroMine may 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 will prolong the analysis time.

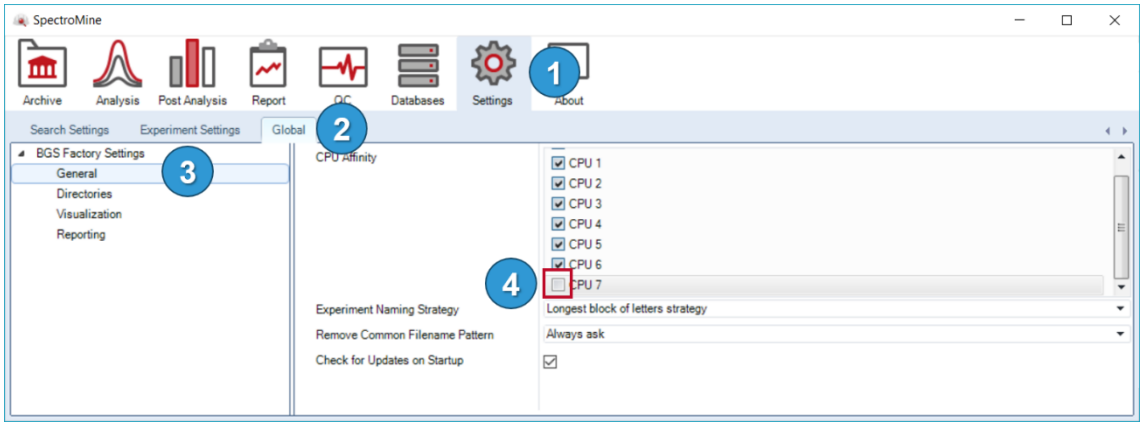

Figure 2. Customize the resources SpectroMine is allowed to use.

1.5 Supported Mass Spectrometers

SpectroMine™ supports mass spectrometers from Thermo Scientific™, Sciex, Bruker and Waters. The specific supported models are:

- Thermo Scientific™ Orbitrap Exploris™ 480
- Thermo Scientific™ Q Exactive™ Series
- Thermo Scientific™ Orbitrap Fusion™ Series
- Sciex TripleTOF® Series (5600, 5600+, 6600)
- Bruker impact II™ and timsTOF and timsTOF Pro
- Waters Xevo® G2-XS QToF, Synapt G2-Si

Box 2. Block Design in ILQ Experiments

In principle, the multyplexing level of an isobaric label quantification (ILQ) experimet is limited by the number of reporter ions. However, thanks to the concept of block design this multiplexity can be increase significantly. The idea is that you divide your samples into blocks, with each block having all channels of an isobaric label tag. The block setup also allows you to define which fractions belong to a sample and summarize them accordingly. Typically, a block contains runs that correspond to the fractions of a sample. If no fractionation was performed, you should assign each run to its own block. For this to work, you need to be able to normalize across blocks afterwards. This is achieved using a common sample labeled with one of the isobaric tags, included in each block, which is normally a pool of all samples. An example of a block design for a TMT 10-plex experiment is shown bellow:

Reporter Ion	Block 1		Block 2		Block 3	
	Condition	*Replicate*	*Condition*	*Replicate*	*Condition*	*Replicate*
126	**Control**	**9**	**Treatment 2**	**5**	**Treatment 2**	**7**
127N	**Treatment 1**	**7**	**Control**	**4**	**Control**	**5**
127C	**Control**	**2**	**Treatment 1**	**5**	**Treatment 1**	**2**
128N	**Treatment 1**	**1**	**Control**	**1**	**Control**	**7**
128C	**Pool**		**Pool**		**Pool**	
129N	**Treatment 2**	**6**	**Treatment 2**	**3**	**Treatment 2**	**1**
129C	**Treatment 2**	**4**	**Treatment 2**	**2**	**Treatment 2**	**9**
130N	**Treatment 1**	**3**	**Control**	**3**	**Control**	**6**
130C	**Treatment 1**	**4**	**Treatment 1**	**8**	**Treatment 1**	**9**
131	**Treatment 2**	**8**	**Treatment 1**	**6**	**Control**	**8**

In this case, thanks to the block design, you can add a significantly higher number of replicates to this study. In a similar way, more conditions could have been analysed.

1.6 Supported Data Acquisition Methods

SpectroMine™ performs two main tasks with MS run files as input: searching, and quantifying isobaric labeled and label-free DDA experiment samples. In general, a reversed phase chromatography is required. We support linear and nonlinear gradients. We recommend a gradient that spans at least 10-35% acetonitrile. We suggest the MS2 scans should cover at least 500-900 m/z of precursor range.

The searching function supports DDA, DIA and PRM data with MS1 information, acquired on one of the supported mass spectrometers.

SpectroMine supports all ion trap resolutions. MS1, MS2, and SPS-MS3 quantification are supported. In addition, neutral loss-based fragmentation label methods (such as EASI-tag) are supported. Fractionation, including gas-phase fractionation, is supported. Block experimental designs are fully supported (Box 2).

DDA data acquired with Ion Mobility annotation such as PASEF and FAIMS are supported for the LFQ, ILQ and labeling efficiency check experiments as well as for library generation. Moreover, SpectroMine supports library generation from diaPASEF, HDMSE and DIA FAIMS data. In order to perform quantitative experiments for the DIA data with ion mobility annotation, we recommend to use Spectronaut - Biognosys software solution for DIA based quantification.

If you have more specific questions about which data we support, or encounter technical problems with the software, please contact us via our Help Center.

2 Getting Started

2.1 Getting SpectroMine™

SpectroMine™ software licenses can be purchased on our webpage. We also provide licenses for a trial period upon request on our webpage. After requesting a license, you will get an email with:

1. **A link to download the installer**
2. **A activation key to activate your copy**

Should you not have received the above information, please contact us via our Help Center.

Important: activation keys are computer-bound and active upon issue. If you need to install SpectroMine on more than one computer, please contact us via Help Center.

2.1.1 SpectroMine Activation

When you install and start SpectroMine for the first time, you will be asked to activate your software by submitting your activation key into the SpectroMine activation dialogue. If your computer has access to the internet, activation will be automatic. If your SpectroMine computer does not have an internet connection, or the connection is blocked by a firewall, 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 (activation.info) on your computer and send this file to us via Help Center. In general, you will receive a license file (license) within one or two working days. To activate SpectroMine using a license file, click on the "Browse License File…" button in the SpectroMine Activation dialogue. **Your license period will start upon its generation (i.e. not upon the activation in the software).**

3 SpectroMine Usage

3.1 Structure of SpectroMine™

3.1.1 Layout

SpectroMine™ is structured in different levels (Figure 3). The highest level is the Perspective. 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, repots and summaries).

Figure 3. SpectroMine general structure. Perspectives are at the top level, tabs with pages take the main space. Data trees with nodes are arranged on the left panel. Main information is shown on the right panels.

3.1.2 Tips for a Better Experience

1. SpectroMine is full of informative text tool-tips throughout the software (Figure 4). They will appear as you hover over many of the elements.

2. Many additional functionalities can be accessed by right-clicking on elements such as experiment tabs, plots, nodes, etc., and intuitive icons to perform main actions such as saving an experiment, reorganizing the tree, revising the experiment settings.

Figure 4. SpectroMine contains additional tips and menus when you hover over some elements, right-click on them, or use the intuitive icons.

3. It is possible you get warning messages like this:

Warnings are sometimes just informational and do not require action. Error messages during library generation or data processing are shown in red. You can display the full 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 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.

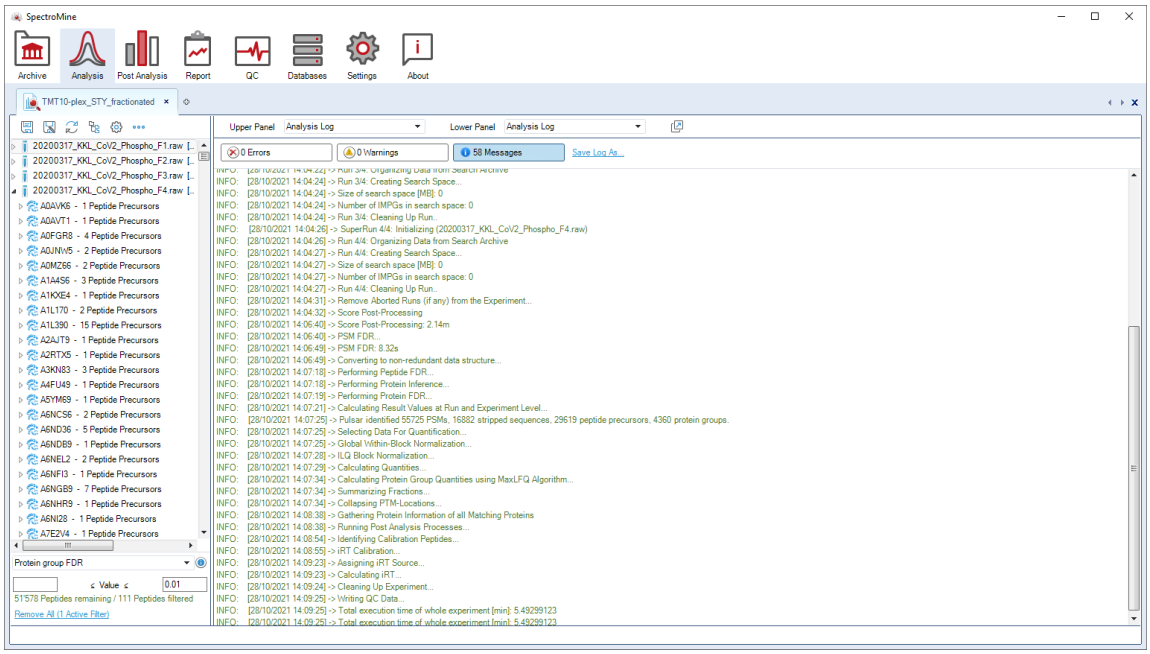

Figure 5. The analysis log shows messages, warnings, and errors in separate tabs. One or more tabs can be selected and simultaneouly displayed.

3.2 Before starting

Make sure you have everything you need ready before starting your analysis in SpectroMine™.

SpectroMine performs different experiments that can be classified into four types :

1. Isobaric label quantitative (ILQ) or label-free quantitative (LFQ) DDA experiments.
2. Labeling efficiency experiments to check labeling ratio and correct annotation of the labels.
3. Database-based search of run files (DDA, DIA and PRM) and spectral library generation from these searches (PSM searches also saved as search archives).
4. Fast and sensitive library-based search of DDA files.

Table 1 shows which resources are required to perform each of the pipelines.

Table 1. Input resources for each task in SpectroMine

Resource	Database search of MS data	Library search of DDA data	ILQ, LFQ	Library generation	Labeling efficiency check
Supported run files	**required**	**required**	**required**	**optional** *(or Search Archives)*	**required**
Protein database (FASTA)	**required**	*not applicable*	**required**	**Optional** *(only for run files)*	**required**
Search Archive	*not applicable*	*not applicable*	*not applicable*	**optional** *(or run files)*	*not applicable*
Spectral library	*not applicable*	**required**	*not applicable*	*not applicable*	*not applicable*
Gene Ontology annotation	**optional**	**optional**	**optional**	**optional**	*not applicable*

3.3 Archive Perspective

The Archive Perspective of SpectroMine™ allows you to review and manage your Search Archives and libraries. In this perspective you can explore basic information about your Search Archives or libraries in their respective tabs. You can also perform actions such as importing and exporting Search Archives and downloading Search Archives from the Biognosys' repository. The most relevant actions are described below.

3.3.1 Search Archives Tree and Overview

In the Search Archives Tree you have your Search Archives grouped by date and then by experiment. If you expand the experiment group, you can see the individual run archives. From this tree, you can download Search Archives from the Biognosys' Search Archives Online Repository. To download an archive, right click and download.

From this panel, you can import and export archives by clicking the corresponding link in the bottom-left corner.

SpectroMine provides several different plots and reports to give you an overview of your Search Archives. You can access these plots by clicking each archive node in the tree and then selecting an item in the right panels (Figure 6). For example, you can get a summary of the search result for the archive, check which settings were used in the search, several bar-plots describing the results, and the Analysis Log of the corresponding search amongst others.

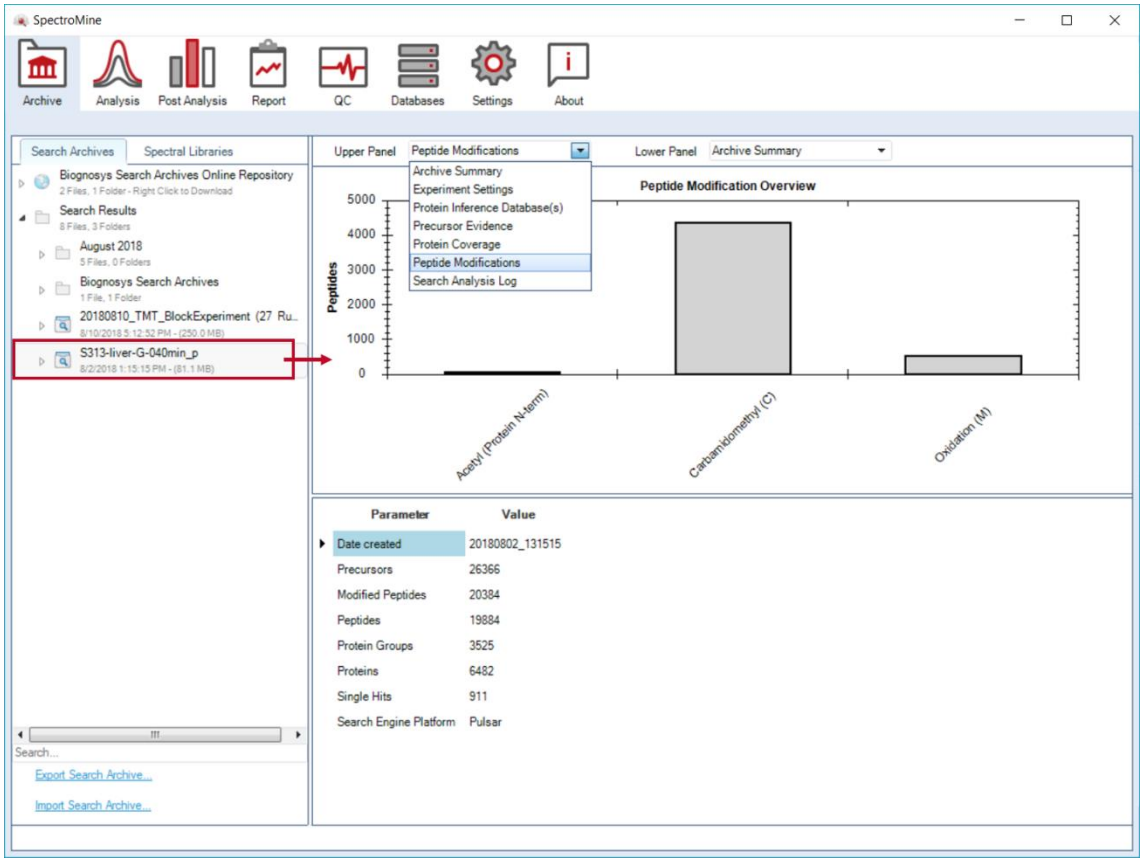

Figure 6. Search Archive overview. Several plots can be selected from the Upper and Lower Panels drop-down menus. In this example, the top plot shows the PTMs overview, while the bottom plot shows the Archive Summary.

3.3.2 Library Tree and Overview

Likewise, for the Search Archives, SpectroMine provides several elements to get general information about your library. You can access these plots by clicking the spectral library node in the tree and then selecting what you want to see in the right panels (Figure 7). For example, you can check general details about what the library contains, several bar-plots depicting your library according to different aspects, information about the protein inference (Zhang *et al.* 2007), or the Analysis Log, among other.You can import and export spectral libraries by clicking the corresponding link in the bottom left corner.

Imported spectral libraries should be in .kit format and corresponding fasta file should be attached.

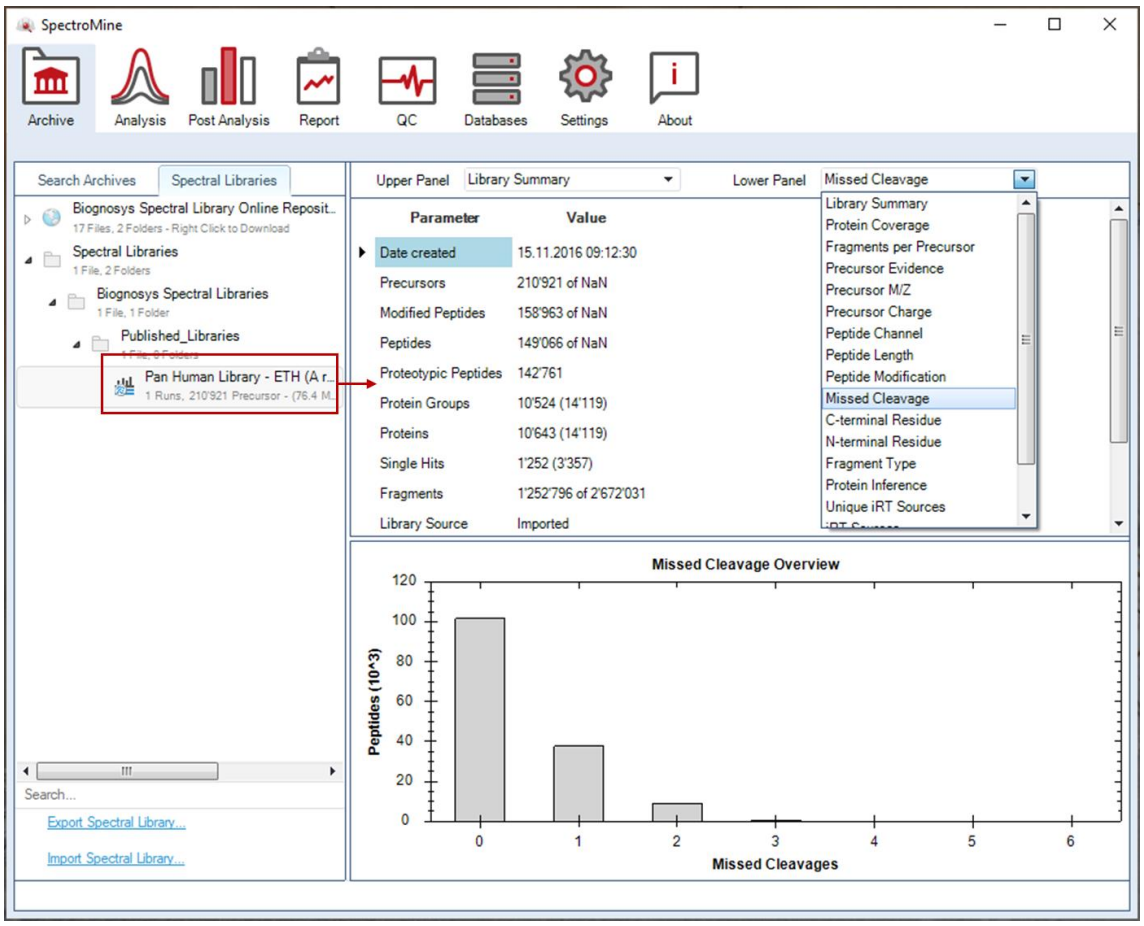

Figure 7. Spectral library overview. Several plots can be selected from the Upper and Lower Panels drop-down menus. In this example, the top plot shows the Library Summary, while the bottom plot shows the missed cleavages overview.

3.4 Analysis Perspective

SpectroMine™ starts up in the Analysis Perspective (Figure 8). This perspective allows you to:

1. Set up and run any SpectroMine Experiment
2. Perform labeling efficiency experiments
3. Generate spectral libraries from raw files and search archives
4. Perform peptide-centric searches of DDA data
5. Review the results of your analyses or searches

Setting up your SpectroMine experiment is easy and straightforward thanks to the set-up wizards. Any type of experiment can be set up following the "Set up a SpectroMine Experiment" link. In addition, the Analysis Perspective provides shortcut links to pre-configured settings for the most usual experiments.

Before starting, see Table 1 to make sure you have everything you need to complete your SpectroMine experiment.

In the next section, we will explain the different steps of the general wizard "Set up a SpectroMine Experiment". However, we strongly suggest using the simplified, preconfigured shortcuts according to the analysis you are aiming to run.

3.4.1 Set up any SpectroMine Experiment

This wizard is very comprehensive and allows you to set up any possible SpectroMine experiment. **We strongly suggest using one of the simplified, preconfigured wizards** if there is one that fits your needs. If this is not the case, go ahead and start the wizard by clicking the corresponding link "Set up a SpectroMine Experiment":

- Give a name to your analysis. Click "Add Runs…" to select the runs files you want to analyze.

- **Choose Fasta Files(s).** In the next window, click "Fasta File" to assign the protein database(s) you want to use for the search. You can select multiple FASTA files, both at the level of experiment and runs.

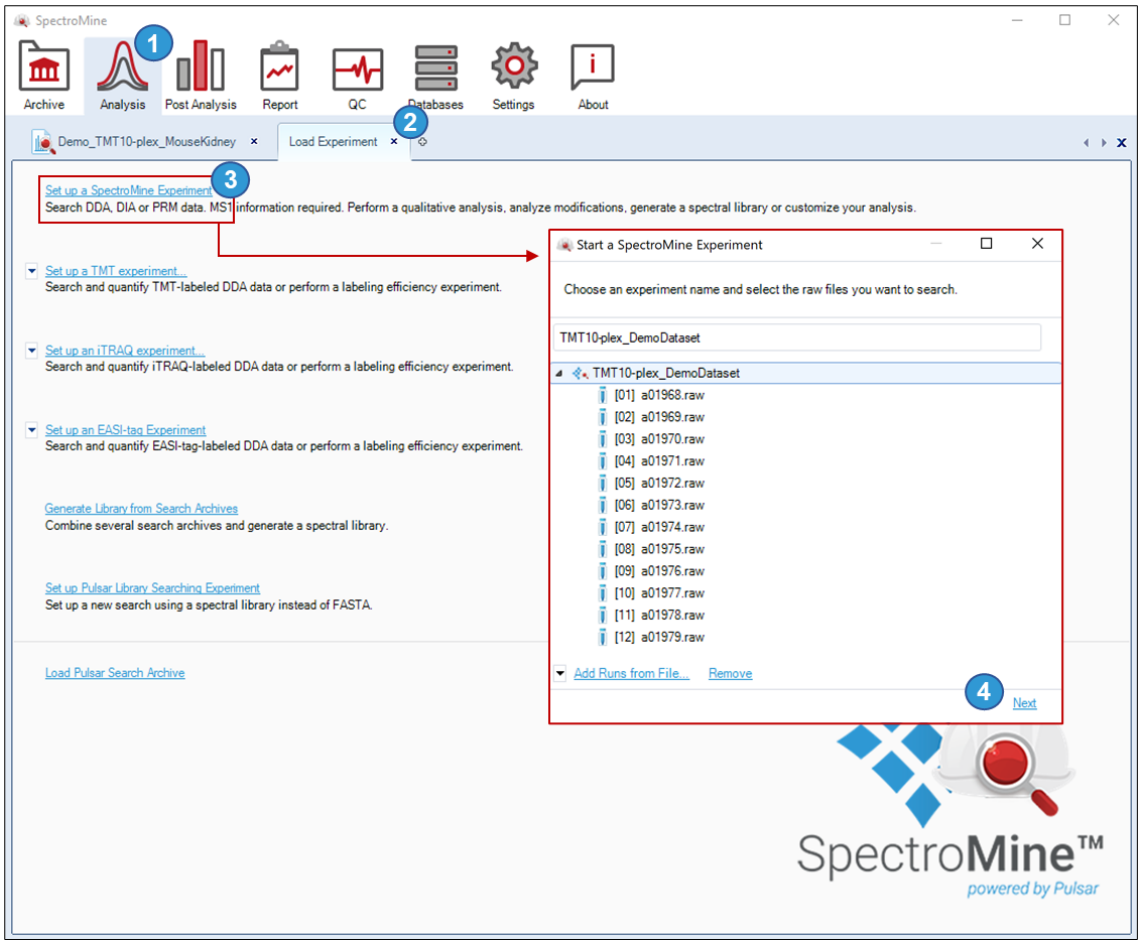

Figure 8. Starting a SpectroMine experiment. After selecting your run files, a wizard will guide you through the process. You will be prompted to assign the FASTA file(s), the GO annotation file, and to select your analysis settings. Finally, you will see a summary of the analysis set-up before clicking "Finish".

- **Choose Pulsar Search Settings.** By clicking "Search Settings…" in the next window you can define your search space (for detailed explanations about each setting, see Appendix 1. Pulsar Search Settings, section 5.1). Choose one of the default schemas, a custom schema or modify one on the fly. Schemas can be assigned at either the experiment or run level. Only one setting-schema can be set per run.
- **If you are generating a library**, you can add Search Archives at this point (for more information about Search Archives, see Box 1). Search Archives prevent you from having to re-search run files if you have already searched them in the past. Select a Search Archive from the tree or import one by clicking "Import Search Archive…".

- In the next window you will be asked to assign runs to blocks. If you don't have a block experimental design, you can ignore this step. SpectroMine will automatically create one block with all your runs.

  If you are running an isobaric label quantification (ILQ) experiment with block design, you need to specify your blocks. To learn what is a block, please refer to Box 2. One block typically contains all fractions of a pooled isobaric labeled sample. This also means a block will typically contain all channels of an isobaric tag.

- ***If you are running an ILQ experiment***, select your quantification method (your tag) in the next window by clicking "Quantification Method…". Otherwise, just click next.

- ***If you are running a comparative analysis***, either ILQ or LFQ, you need to define your experimental set-up conditions. First, in the next window, define your conditions (specify a label in ILQ) and choose colors for plotting purposes (Figure 13).

- In the next window you can annotate your runs (or your channels in ILQ) with information about the condition and the replicate they correspond to (Figure 14). At this point, if you are analyzing isobaric labeled samples, using block experiment, you can select which channel should be used for cross-block normalization (see more in section 3.4.2).

  Based on this information, SpectroMine will test for differentially abundant proteins.

- In the next window, you have the possibility of performing protein inference on a different FASTA file than the one used for searching. SpectroMine performs protein inference according to the IDPicker algorithm (Zhang *et al.* 2007). Choose one of the FASTA files available in the tree or click "Import…" to add one from file.

- In SpectroMine you can add Gene Ontology (GO) annotation information to your experiment. This will provide further biological insight to your results. To select a GO annotation file (*.gaf) at this point, you need to have the file already imported in the Databases Perspective. Learn how to do so in section 3.8.5.2.

- Finally, choose your experiment settings. Your experiment will be saved in the shape of a Search Archive if you check the "Generate a Pulsar Search Archive" check-box. Specify your quantitative settings, your FDR thresholds, etc. In addition, if you are generating a library, specify your Library Generation Settings at this point. Find a

detailed explanation of the Experiment Settings in Appendix 2. Experiment Settings (section 5.2).

After completing the wizard and clicking "Finish", SpectroMine will switch back to the Analysis Perspective and start the analysis. You can monitor the status of the process thanks to the bar and text hints in the bottom-left corner.

In addition to the wizard described above, SpectroMine provides simplified, preconfigured shortcuts for the most frequent experiments (Figure 9). See below for detailed explanation of each of these wizards.

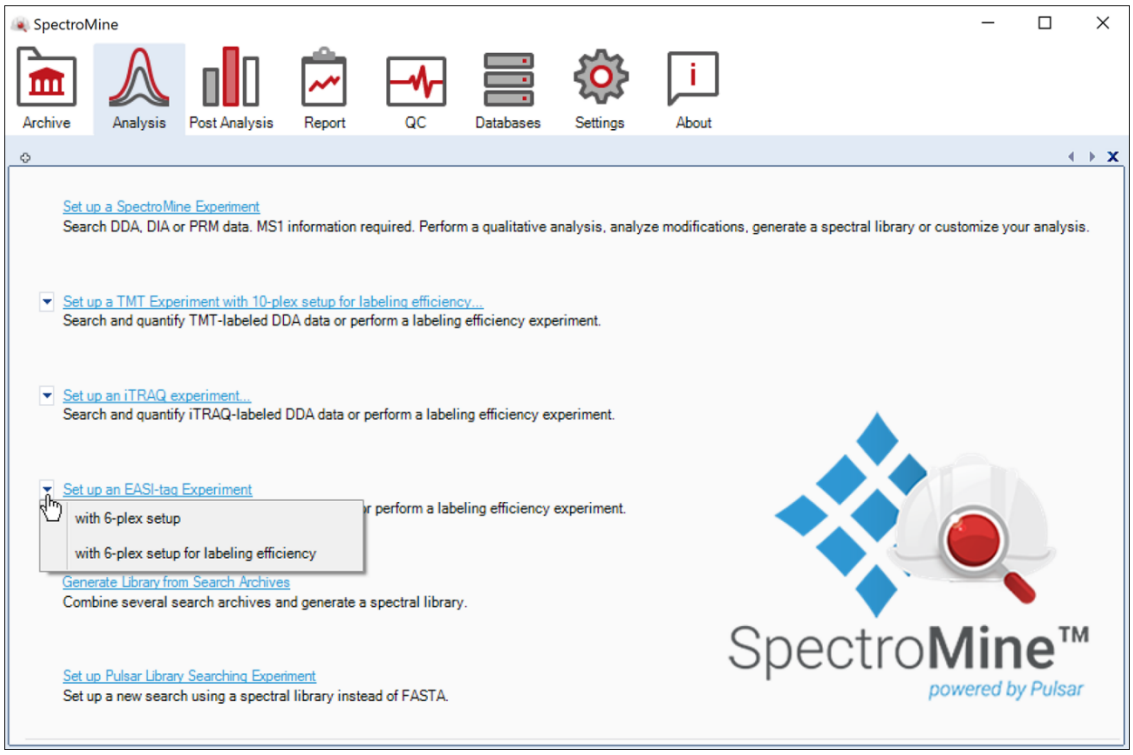

Figure 9. Link to different actions in SpectroMine. Each of these links will guide you through a wizard with pre-configured settings for each workflow.

3.4.2 Set up an ILQ Experiment

SpectroMine supports and has preconfigured wizards for ILQ experiments using TMT tags (6-plex, 10-plex, 11-plex, pro 16-plex or pro 18-plex), iTRAQ (4-plex and 8-plex), and EASI-tag (6-plex only, currently). Use the appropriate shortcut in the Analysis perspective to start a preconfigured wizard and proceed as described in section 3.4.1.

After naming the analysis, selecting the runs files, and the FASTA file for the search, SpectroMine will automatically load the default search setting schema based on the chosen workflow. The “TMT_Quantification” schema can be used for 6-plex, 10-plex, and 11-plex since the tags are the same, whereas the “TMTpro_Quantification” schema is made for TMT16-plex or TMT18-plex. The search schema also specfifies other paratemers that can be modified: the digestion type, the fragment ion type, if quantification is performed at the MS3 level, if speed-up mechanisms (e.g. MS2 Indexing) should be used. For detailed explanations of each setting, see Appendix 1. Pulsar Search Settings, section 5.1.

In the next window, you will be asked to assign runs to blocks. If you don't have an experimental block design, you can ignore this step; SpectroMine will automatically create one block with all your runs Box 2. One block typically contains all fractions of a pooled isobaric labeled sample. This also means a block will typically contain all channels of the chosen isobaric tag. Learn more about block design to increase multiplexing in Box 2. Figure 10 explains how to assign runs to blocks in SpectroMine:

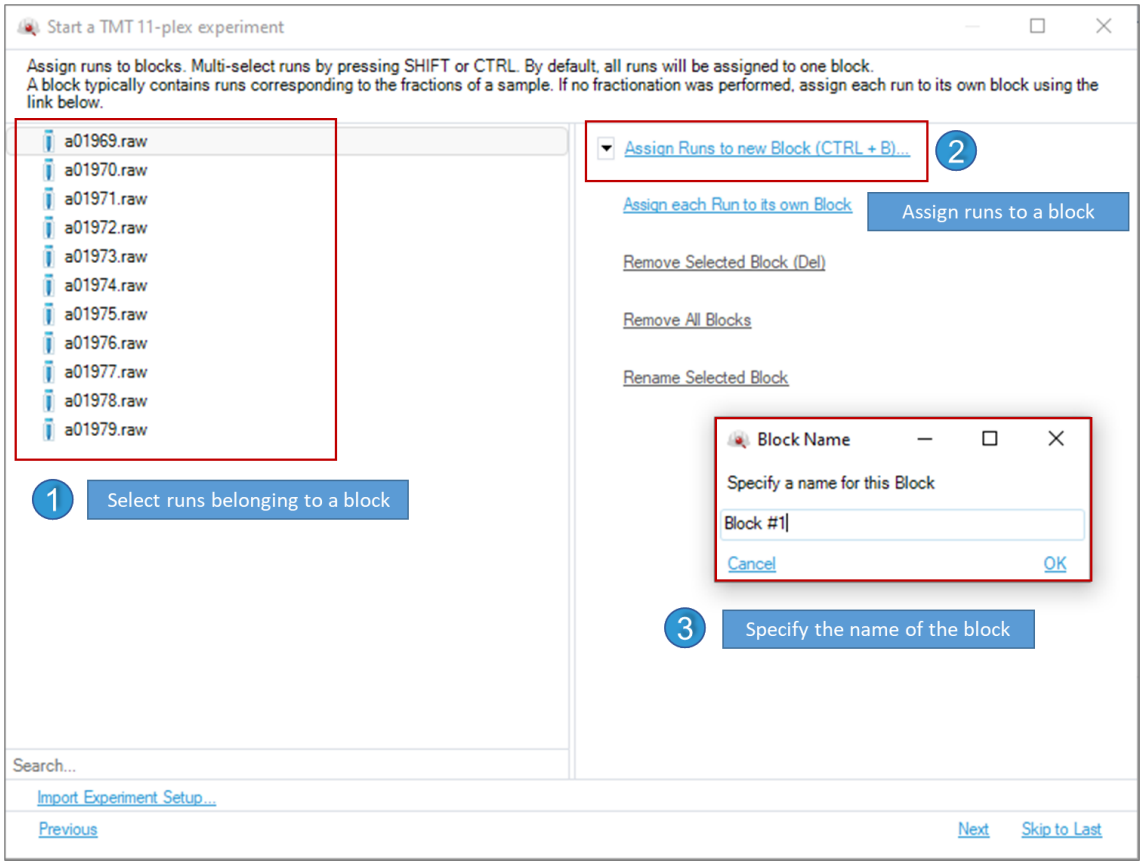

Figure 10. SpectroMine supports experimental block design. Blocks can either be manually specified or imported as previously saved Experiment Setup (yellow circle).

Using the option “Assign each Run to its own Block” you can also assign each of the runs to separate blocks. This option should be used if separate runs are not fractions of the same sample but rather different replicates and/or experiemntal conditions. If all runs were assigned to one block but should be assigned to separate ones, you can quickly remove assigned blocked from all of them by choosing an option “Remove all Blocks”.

In the next window, SpectroMine will automatically select the pre-configured method used for quantification at the MS2 or MS3 level (according to what specified in the search settings). When specifying a quantification method, make sure that the channel tags apply to the labels (fixed modifications) previously selected for the search (Figure 11).

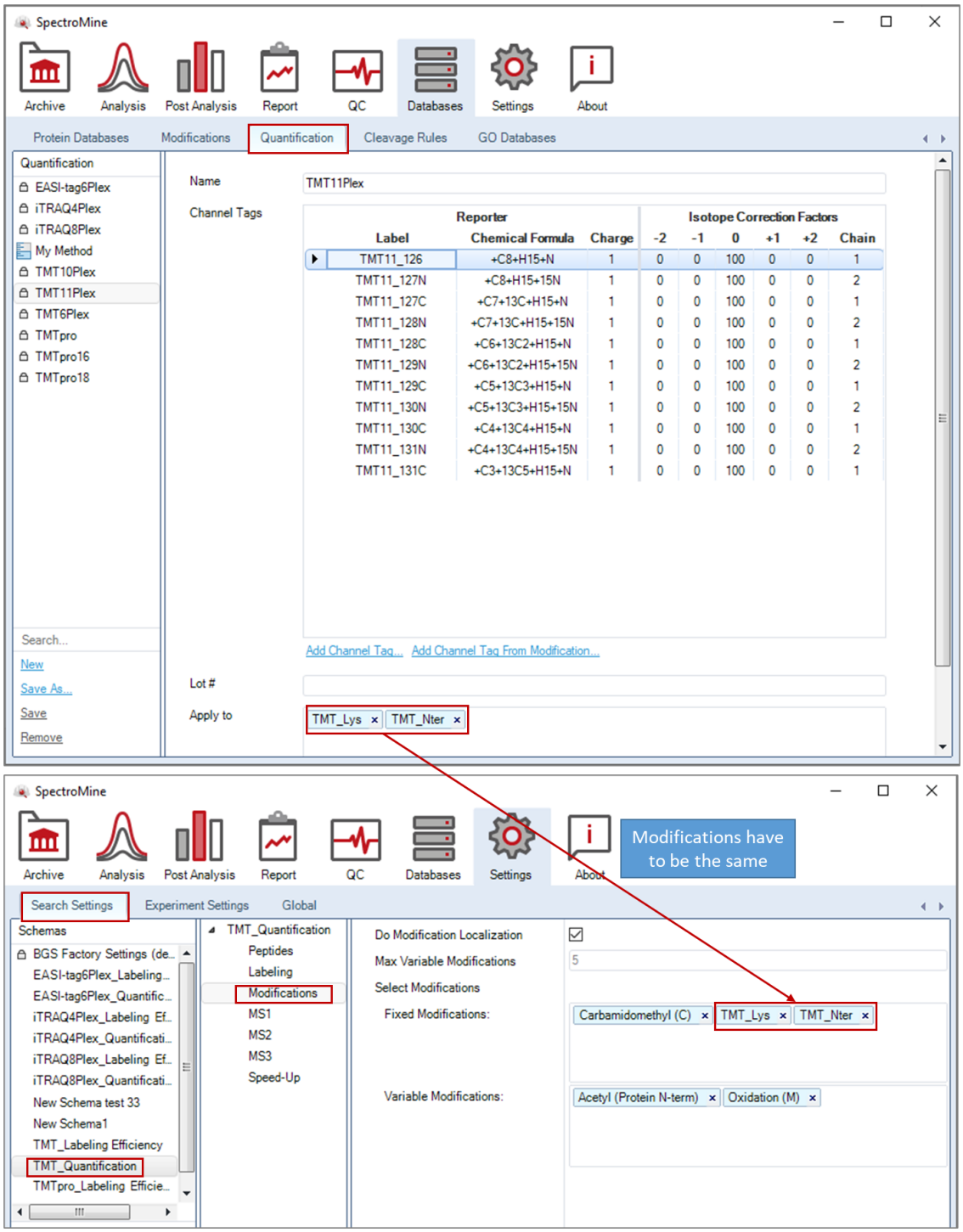

Figure 11. When setting up an ILQ analysis, confirm that labels (fixed modifications), channel tags (e.g. TMT10-plex), and quantification level (MS2 or MS3) are correctly set.

The reporters' label, chemical formula, charge as well as the isotope correction factors specified in the quantification method are provided by the supplier of the isobaric label batch (Figure 12). Labels can be peaks (e.g. TMT) or losses (e.g. EASI-tag, Box 3, page 32). In the quantification method, the column "Charge" specifies whether a reporter ion

is a peak or a loss. The “Chain” specifies, which channels are connected when it comes to impurities.

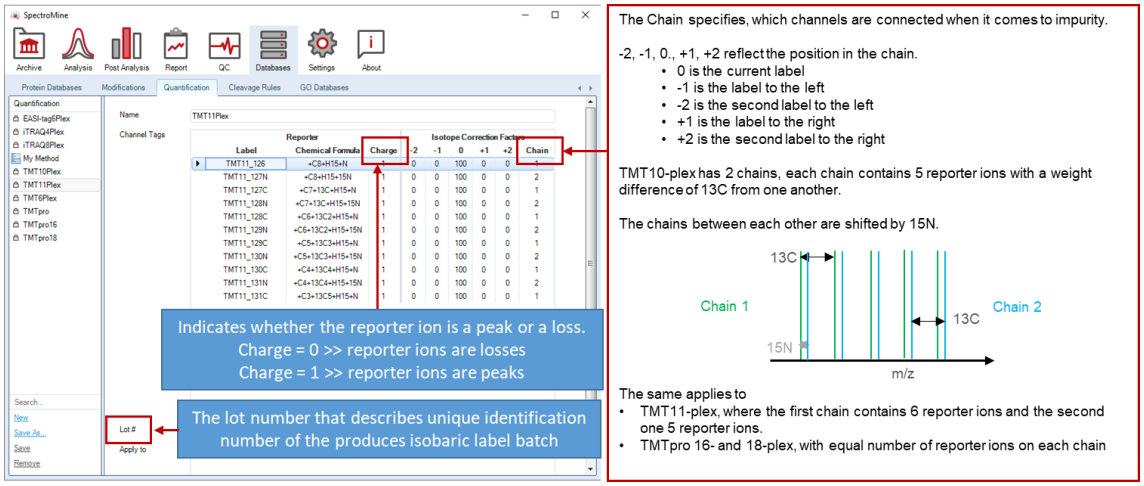

Figure 12. Based on the isobaric label batch, the supplier provides instructions on the reporter ions (Charge) and cross-channel impurities (Chain).

In the next window, you can define the experimental conditions to perform statistical testing in the post-analysis (see section 3.5.8). The conditions are named and colors are chosen for plotting purposes (Figure 13). Only if the experimental conditions are assigned, the next window allows to specify conditions per block and to connect conditions to channels (Figure 14). At this point, if you have a block experimental design, you should also select which channel should be used for cross-block normalization. There is the option of exporting the experiment setup in tab-separated format for future experiments.

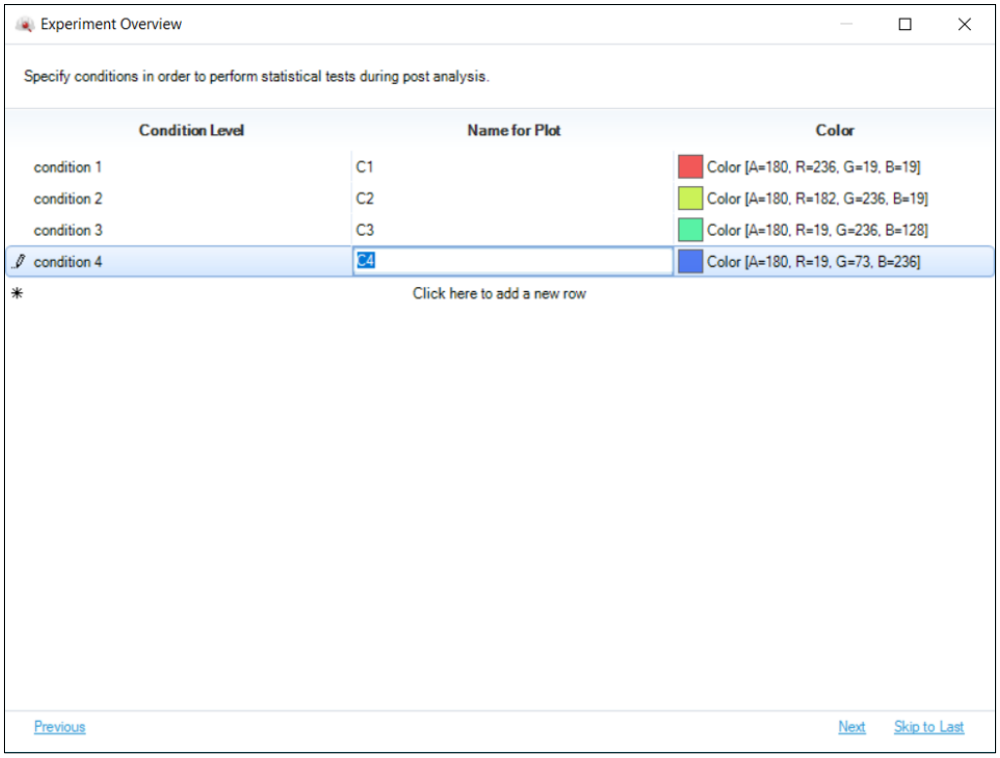

Figure 13. Specify the different conditions in your experiment for differential abundance testing. Choose names and colors for post-analysis plots.

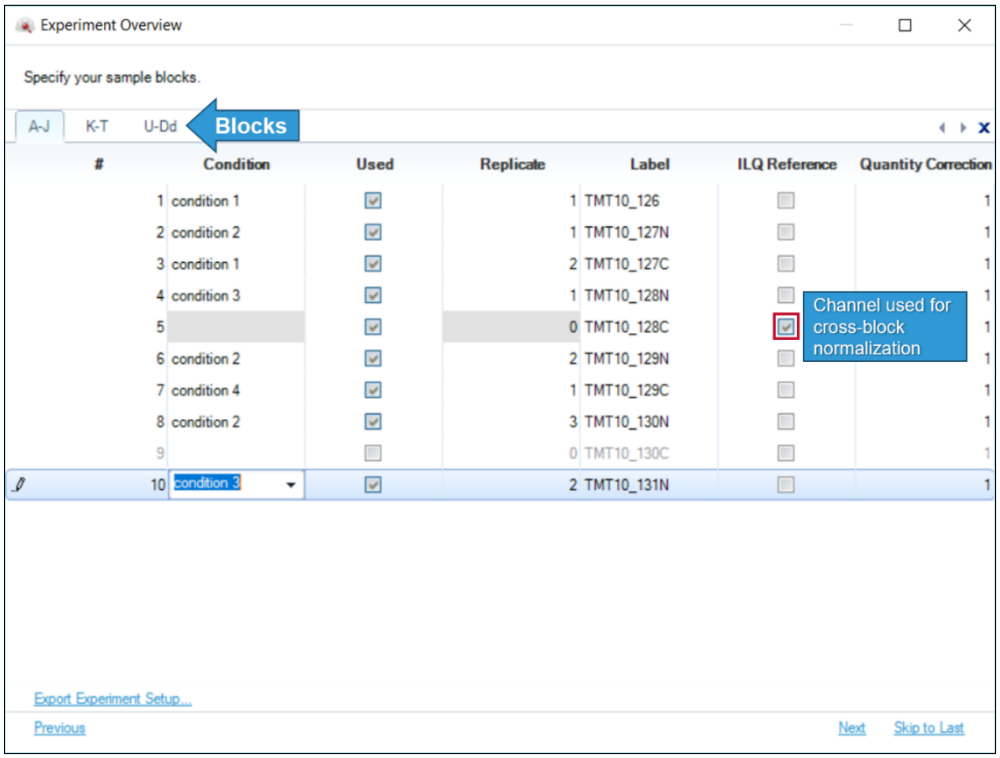

Figure 14. In an ILQ experiment, specify to which condition and replicate each channel corresponds. If you have a block design with cross-block normalization, specify which channel should be use for such purpose.

You can add Gene Ontology (GO) annotation information to your experiment. This will provide further biological insight to your results. To select a GO annotation file (*.gaf) at this point, you need to have the file already imported in the Databases Perspective. Learn how to do so in section 3.8.5.2.

Next, choose your experiment settings (Figure 15). The Quantification schema is automatically selected for ILQ preconfigured shortcuts (Figure 11, top panel). Some important parameters to define when processing ILQ experiemnts on Thermo data are the ILQ filters and missing value imputation. The Peak-to-Threshold filter says that when the intensity of a precursor ion divided by the noise is lower than a set value, its PSMs are not used for quantification. The Signal-to-Interference filter says that when the summed signal intensity of a precursor ion divided by the summed intensity of the whole precursor selection window is lower than a set value, its PSMs are not used for quantification. The ILQ missing value imputation replaces a missing quantity by a random value between 0 and the per-scan noise information provided by the raw data (default: enabled).

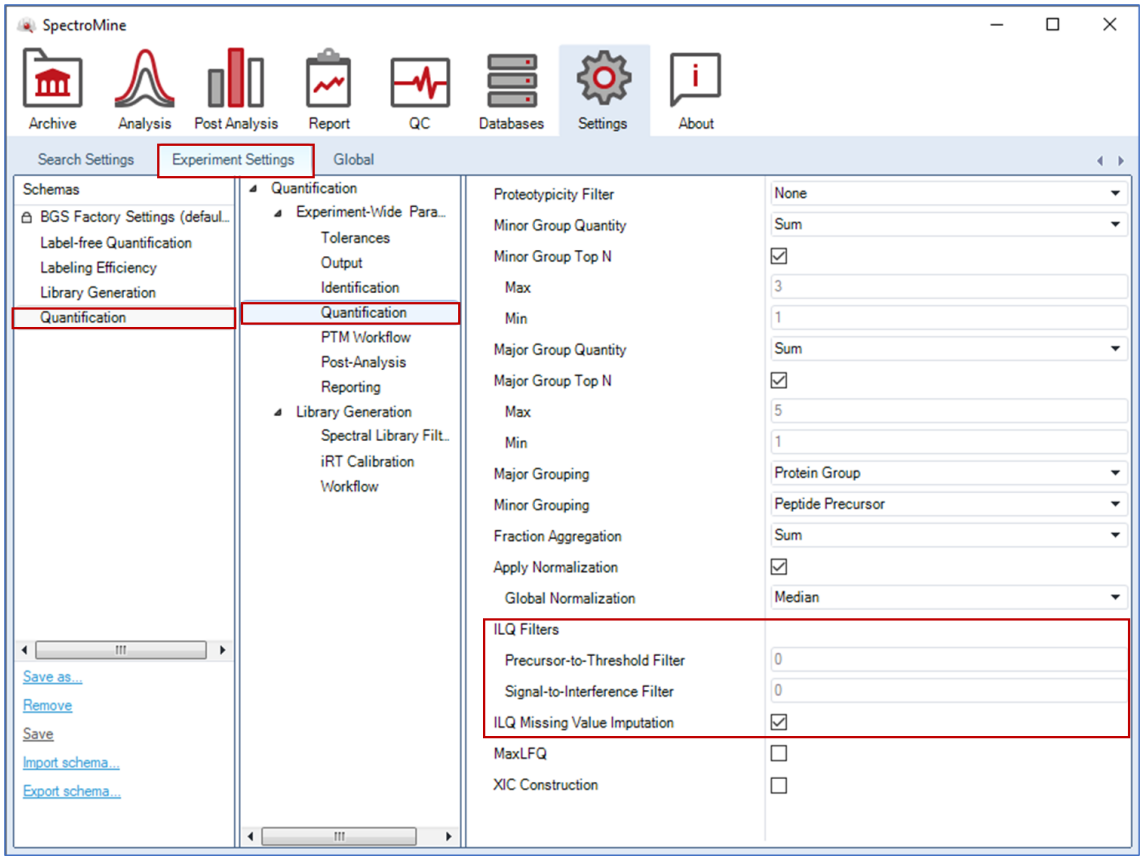

Figure 15. The Quantification schema in the Experiment Settings is used for ILQ. Every parameter included in the section «Experiment-wide parameters» apply to ILQ experiments. The «Library Generation» pipeline is not executed for quantification.

In the Post-Analysis node you can set the parameters for differential abundance testing. By default differential abundance will be tested on protein group level (observation groups). All quantities on peptide precursor level will be used for the statistical testing (observations). After revising your analysis in the summary window, you can press "Finish". SpectroMine will switch back to the Analysis Perspective and start the analysis. You can monitor the status of the process thanks to the bar and text hints in the bottom-left corner.

Box 3. Easi-tag Quantification Concept

For the quantification using EASI-tag, SpectroMine includes information from precursor and fragment ions. The low m/z of EASI-tag's reporter ion is equivalent to a neutral loss, which typically also the b- and y-ions lose (Virreira Winter *et al.* 2018).

Therefore, for each PSM SpectroMine first filters for quantification relevant ions and clusters them according to their ion type and position. After applying isotope correction to the clustered ions, the TopN clusters are selected for summarization. In the end for each channel SpectroMine returns 1 summarized and isotope corrected quantity.

3.4.3 PTM Workflow

SpectroMine features PTM workflow dedicated to determination of posttranslational modifications occurrence, confidence of their localization and their differential abundance analysis.

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. 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 16). 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 16 Example of quantitative site collapse of phosphorylated parent peptides, performed according to their multiplicity.

3.4.4 Set up a Labeling Efficiency Check Experiment

The purpose of the Labeling Efficiency Check (LEC) Experiment in ILQ is to verify if:

1. The workflow results in a high percentage of labeled data
2. The annotation of the labels is correct

To address these two concerns, a LEC experiment counts:

1. The labeled vs. unlabeled PSMs
2. The number of reporters per channel

If you want to run a Labeling Efficiency Check (LEC) Experiment, there are various preconfigured wizards for that purpose in the startup page of the Analysis Perspective. Each of the preconfigured ILQ experiments has a corresponding LEC Experiment wizard, which can be found in the drop down menu of each isobaric label (Figure 17). Additionally, you can set up any LEC Experiment of your choice using the general set up wizard by clicking the "Set up a SpectroMine Experiment" link.

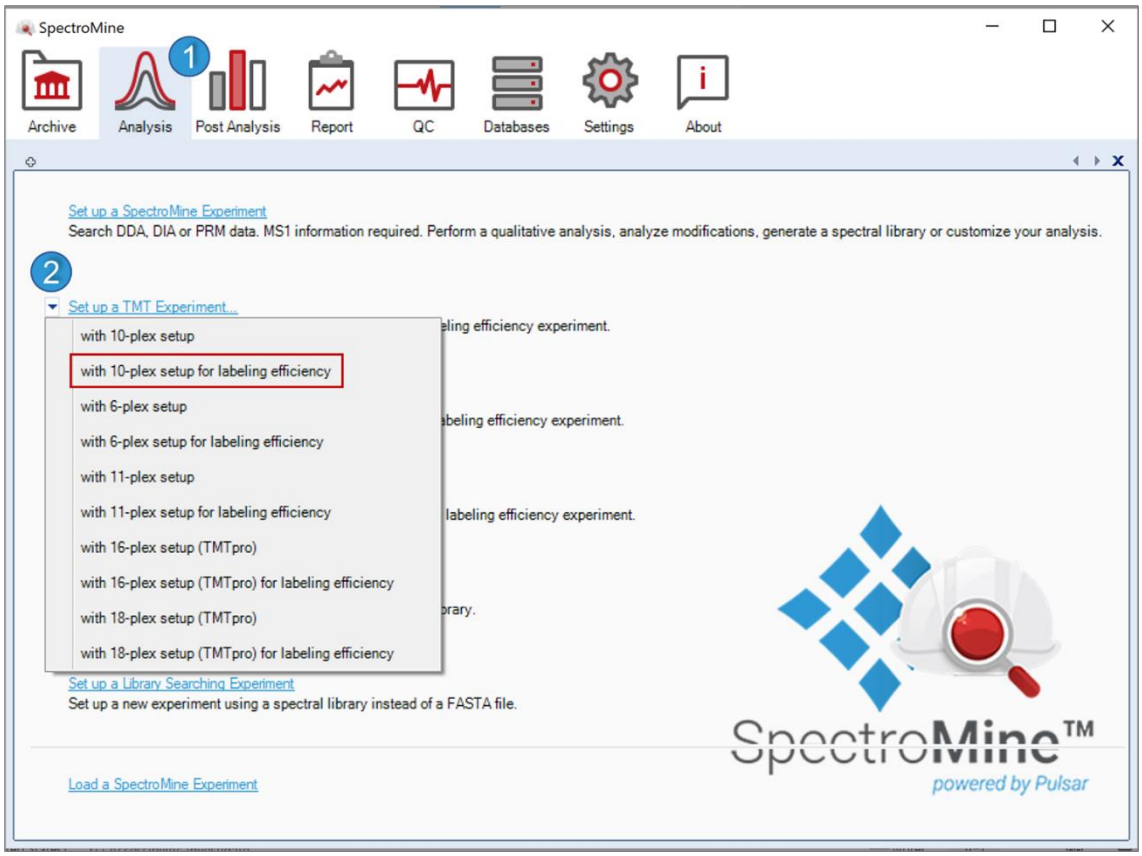

Figure 17 Links to different LEC Experiment wizards in SpectroMine.

Before setting up a non preconfigured LEC Experiment, you first have to configure

- a Quantification Method with the labels of your choice (for detailed explanations how to setup a Quantification Method see 3.8.3 Quantification),
- Search Settings, where the ILQ labels are set as variable modifications and are equal to the ones configured in the Quantification Method. (Figure 18), and

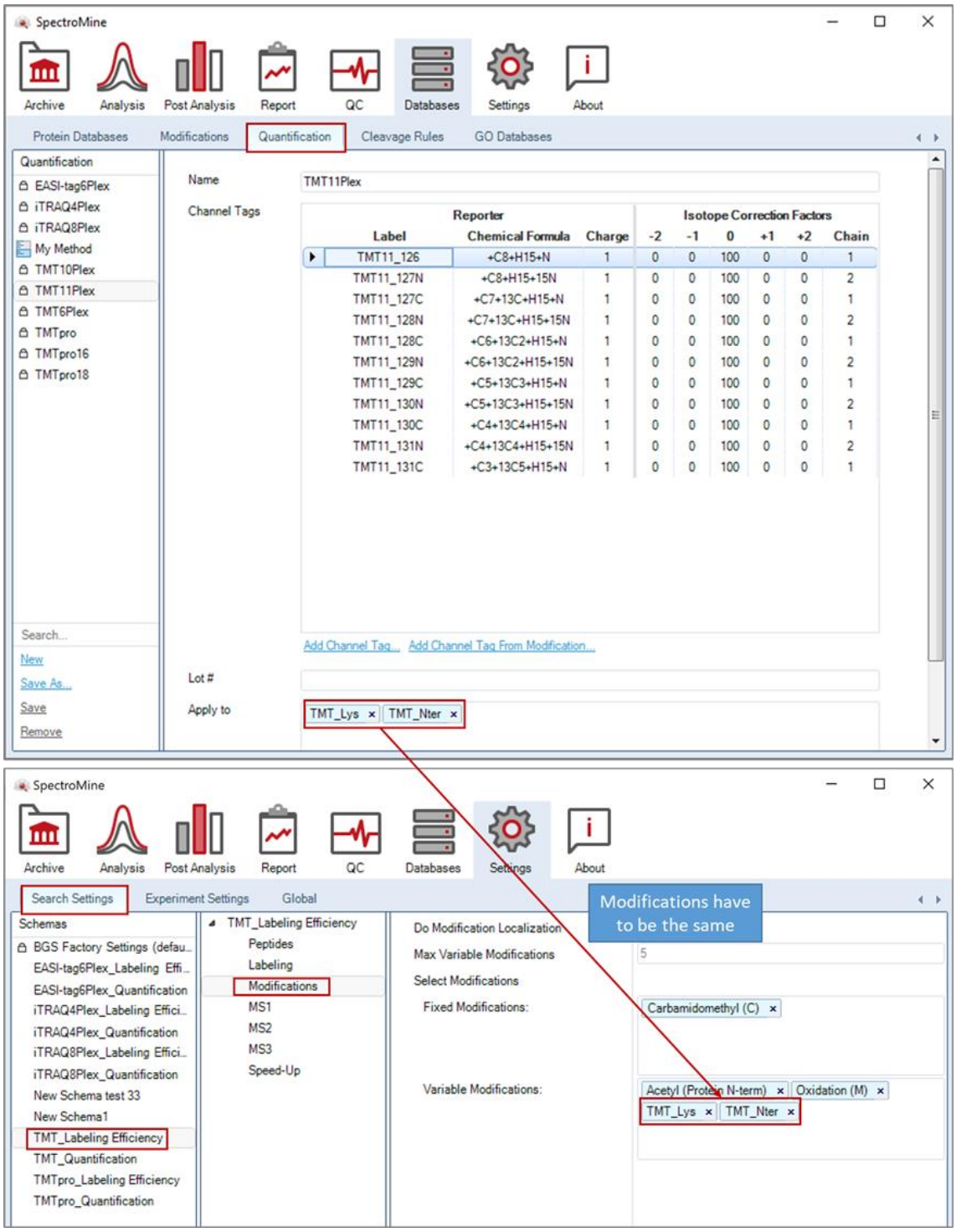

Figure 18 TMT10plex example of required configuration for a LEC Experiment. (a) Configured Quantification Method. (b) Configured Search Settings.

- Experiment Settings with selected “ILQ Efficiency” feature to apply a LEC Experiment (Figure 19).

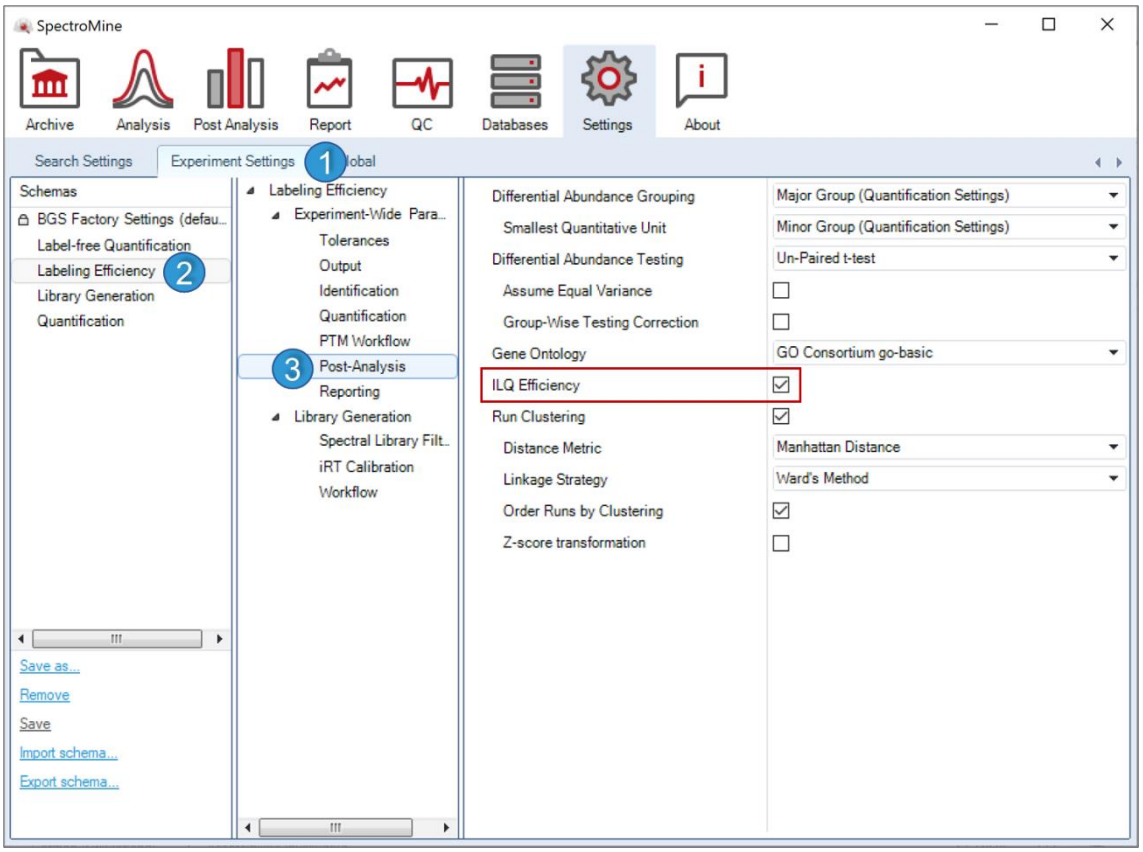

Figure 19 Experiment settings of LEC Experiment

All these configurations are taken care of in the preconfigured LEC Experiment wizards.

The wizard will start at this point to set up a preconfigured LEC Experiment

- Give a name to your analysis. Click "Add Runs…" to select the run files you want to analyze.

- **Choose Fasta Files(s).** In the next window, click "Fasta File" to assign the protein database(s) you want to use for the search. You can select several FASTA files, and you can do that on experiment or on run level.

- SpectroMine will automatically load the corresponding LEC Search Settings in the next window. However, you can click "Search Settings…" to modify aspects of the search space (for detailed explanations about each setting, see section 5.1 Appendix 1. Pulsar Search Settings).

- SpectroMine will automatically load the corresponding LEC Quantification Method in the next window. However, you can click “Quantification Method…” to select a different one. However, it is important that is fits to the Search Settings as depicted in Figure 18.
- SpectroMine will automatically load the corresponding LEC Experiment Settings in the next window. However, you can modify the specified quantitative settings, your FDR thresholds, etc.

After completing the wizard and clicking "Finish", SpectroMine will switch back to the Analysis Perspective and start the analysis. You can monitor the status of the process thanks to the bar and text hints in the bottom left corner.

3.4.5 Generate Library from Search Archives

If you want to generate a library from Search Archives, click the "Generate Library from Search Archives" link in the start-up page of the Analysis Perspective.

- Give a name to your library.
- Select the Search Archives (for more information see Box 1) from the tree or import one by clicking "Import Search Archive…".
- In the next window, you have the possibility of performing protein inference (Zhang *et al.* 2007) on a different FASTA file than the one used for searching. Choose one of the FASTA files available in the tree or click "Import…" to add one from file.
- In SpectroMine you can add Gene Ontology (GO) annotation information to your library. This will provide further biological insight to your results. To select a GO annotation file (*.gaf) at this point, you need to have the file already imported in the Databases Perspective. Learn how to do so in section 3.8.5.2.
- Finally, choose your experiment settings. Specify your Library Generation Settings at this point. Find a detailed explanation of the Experiment Settings in Appendix 2. Experiment Settings (section 5.2).

After completing the wizard and clicking "Finish", SpectroMine will switch back to the Analysis Perspective and start the analysis. You can monitor the status of the process thanks to the bar and text hints in the bottom-left corner.

3.4.6 Perform a Library-Based Search

It is also possible in SpectroMine to do a search using a spectral library instead of a FASTA file to generate the search space. Both types of searches have a spectrum-centric approach, where each spectrum is queried against a list of peptides (unlike a peptide-centric approach, where each peptide is queried against spectra). The only difference between them is how the search space is generated. In case of FASTA-based search, digest and modification settings are selected to in silico generate all the peptides of interest. In a library-based search, this is not necessary as only the peptides from the library are used. The advantages of using a library-based search over a FASTA-based search is that the search space will be much smaller, making the search a lot faster. On top of that, a library contains a lot of prior information (peptides that were identified before, fragmentation patterns, …) that will improve the sensitivity of the search. To perform a library-based search, click on the link “Set up Pulsar Library Searching Experiment”, and the wizard will open:

- **Give a name to your analysis.** Click "Add Runs…" to select the runs files you want to analyze.

- **Choose Library(/ies).** In the next window, select all the libraries you want to use for the search.

- **Choose Search Settings.** Some settings (e.g. digest and modifications settings) are greyed out, as the library contains that information already.

- **Condition Setup.** It is possible to do a labeled (if the library has the modifications of interest) or label-free quantitative experiment. In that case, specify blocks, quantification method (in case of ILQ), conditions and condition setup.

- **Gene Annotation.** In SpectroMine you can add Gene Ontology (GO) annotation information to your experiment. This will provide further biological insight to your results. To select a GO annotation file (*.gaf) at this point, you need to have the file already imported in the Databases Perspective. Learn how to do so in section 3.8.5.2.

- **Experiment Settings.** Finally, choose your experiment settings. Your experiment will be saved in the shape of a Search Archive if you check the "Generate a Pulsar Search Archive" check-box. Specify your quantitative settings, your FDR thresholds, etc.

After completing the wizard and clicking "Finish", SpectroMine will switch back to the Analysis Perspective and start the analysis. You can monitor the status of the process thanks to the bar and text hints in the bottom-left corner.

3.4.7 Reviewing your analysis

An analysis is shown in the Analysis Perspective organized as a data tree on the left and plots, reports and summaries on the right side (Figure 20). By default, the hierarchy of the data tree is:

>Run
  >Protein Group
    >Peptide Precursor only present for the isotopically labeled samples
      >Precursor (m/z)
        >PSM

You can change the tree structure by right-clicking on the experiment tab → Group by, then select one of the options. The most common actions are also accessible through intuitive icons that are available under the experiment tab.

The right side is divided into two panels (Upper and Lower panels), so you can look at two different plots at the same time. These plots will change based on what is selected in the data tree. Different plots are applicable to different levels. To know which plots are available for each level and their meaning, see Appendix 3. Analysis Perspective Plots (section 5.3).

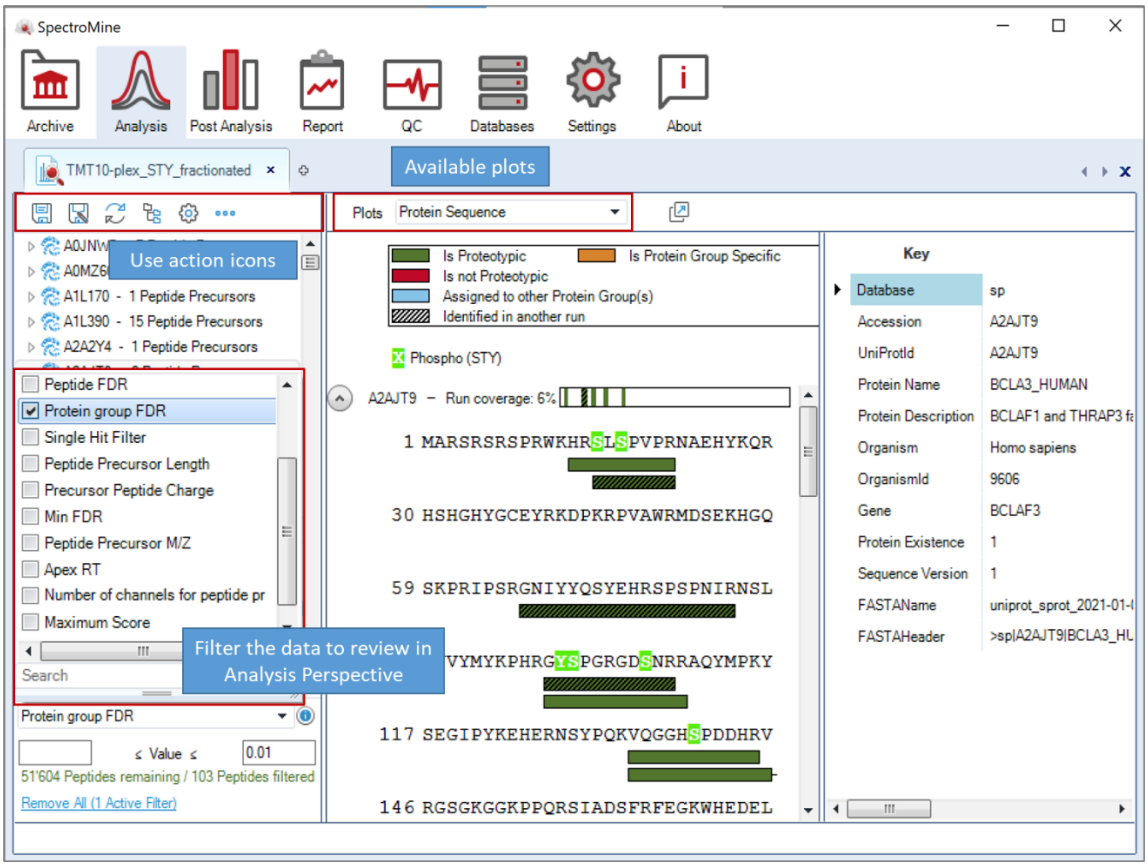

Figure 20. Analysis view in the Analysis Perspective. Data tree is on the left side. Plots and summaries are on the right side. A summary of the number of identifications is shown in green at the bottom. Filtering the tree in the Analysis Perspective. Check the box for a filter and give the corresponding value.

3.4.7.1 Analysis Perspective Plots

One of the main tools SpectroMine provides for analysis reviewing in the Analysis Perspective is the set of plots and reports available for the different levels of the data: run, protein group, peptide precursor, and peptide-spectrum match (PSM).

- Run level reports: overview of the iRT calibration (Box 3), summary overview of the run, mass calibration, TICs, score histograms, PSMs per search, and log of the analysis.
- Protein, precursor and PSM level plots: protein inference and protein coverage information with PTMs site locations, PSM plot with visualization of diagnostic peaks, XIC of MS1 isotope envelop, MS3 spectrum, if applicable. Please, visit the Appendix 3. Analysis Perspective Plots (Section 5.3). to find an example and a description of each plot.

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

Box 3. iRT Calibration in SpectroMine

SpectroMine will perform the best, most accurate possible iRT calibration on each run file (Escher *et al.* 2012; Bruderer *et al.* 2016). For this purpose, SpectroMine will apply Deep Learning Assisted iRT Regression. The new Deep Learning algorithm will be used to generate the iRT reference set, which is useful when working with non-model organisms hardly covered in Spectronaut's internal empirical iRT reference dataset. Otherwise, the user can choose calibration based on that empirical iRT Database. SpectroMine contains a highly comprenhesive repository of commonly found peptides from several organisms that are used as iRT peptides. By default, SpectroMine will try to perform the non-linear –precision- iRT calibration. If the number of datapoints is not enough to do that, it will try to apply a linear iRT calibration. In case this is still not posible, due to the fact that the run is of very low complexity, comes from a very uncommon organism, or similar, SpectroMine will apply an *in silico* iRT calibration. In this case, we highly recommned to spike the iRT Kit into your samples, so at least the linear calibration can be performed.

3.4.7.2 Tree Filtering

Using the filtering system implemented in SpectroMine, 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 20) 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 convey with all selected filters in order to be shown in the review tree. By default, the precursor and protein Q-value filters are set to what was chosen in the settings (default is 0.01).

Note: sometimes it is nonobvious that a filter is applied. Make sure you don't have any unwanted filter applied before reviewing your analysis further.

3.4.8 Experiment Tab Actions

In the Analysis Perspective, right-click on the experiment tab. A context menu will open with several functionalities to apply to the analysis **Figure 21**. The detailed descriptions of experimental tab actions are located in list below:

Map missing runs	If SpectroMine lost the link to the run files, you can map them back.

Refresh Post Analysis	Apply changes you submitted in the settings.
Save As…	Save your analysis in Search Archive format (*.psar).
Export All… (Ctrl + R)	With this function, you can batch export many reports of your choice relevant to your analysis. A window will open to let you choose which reports and logs you want to save. You can choose your destination here.
Group By	You can group your data tree under different criteria: All Precursors, Modified Sequence, Stripped Sequence or Protein Groups (default setting).
Settings	This tool allows to explore and **EDIT** several aspects of your analysis: • Your Experiment Overview (see section 3.4.6) • The FASTA file for protein inference (Zhang *et al.* 2007) • The Gene Ontology annotation file (*.gaf) to retrieve biological insight to your data • Quantification, FDR cut-offs and Post Analysis settings Changing these settings will let you recalculate your analysis, which is significantly less expensive than running it again from scratch.

The most common actions are also accessible through intuitive icons.

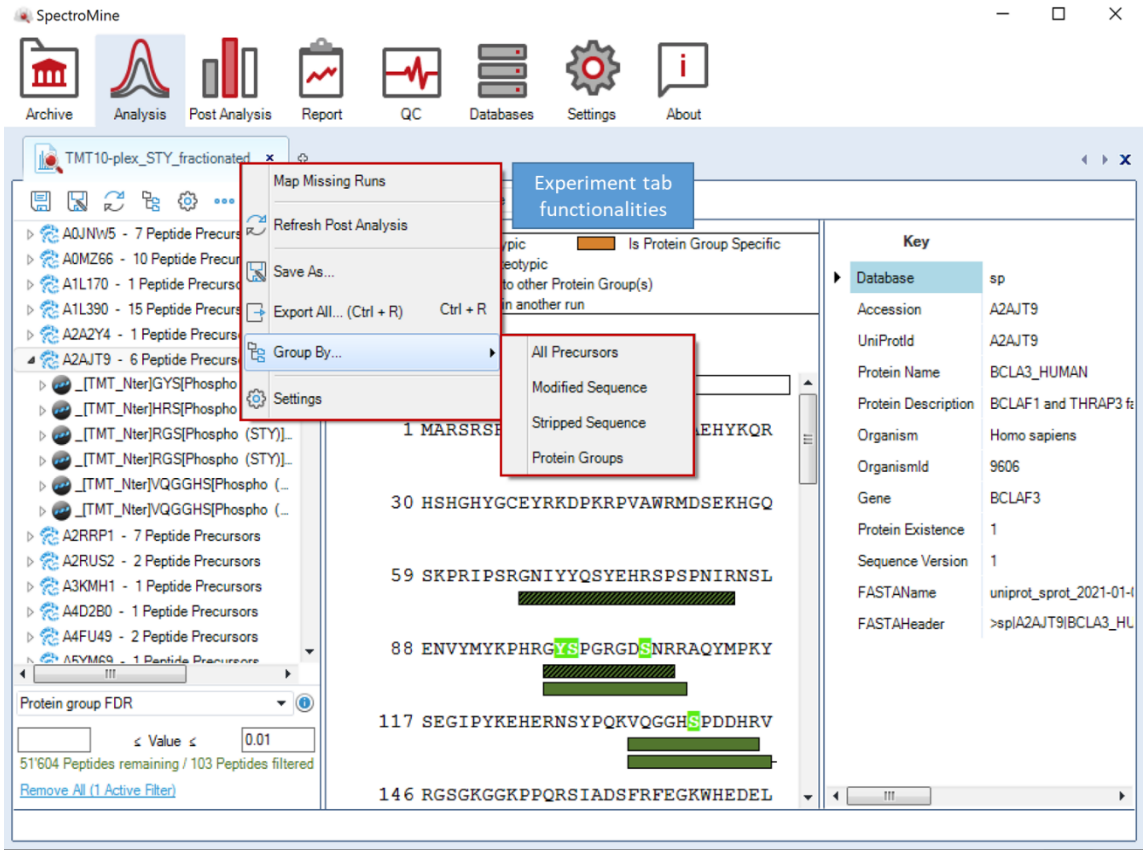

Figure 21. Right clicking the experiment tab will allow you to run some functionalities like save the experiment and change the settings of the analysis.

3.5 Post Analysis Perspective

The Post Analysis Perspective in SpectroMine™ reports information that is less raw data focused as compared to the Analysis Perspective. It shows summary information about identification, quantification, results of the differential abundance test, hierarchical clustering, and GO terms enrichment & clustering (Figure 22).

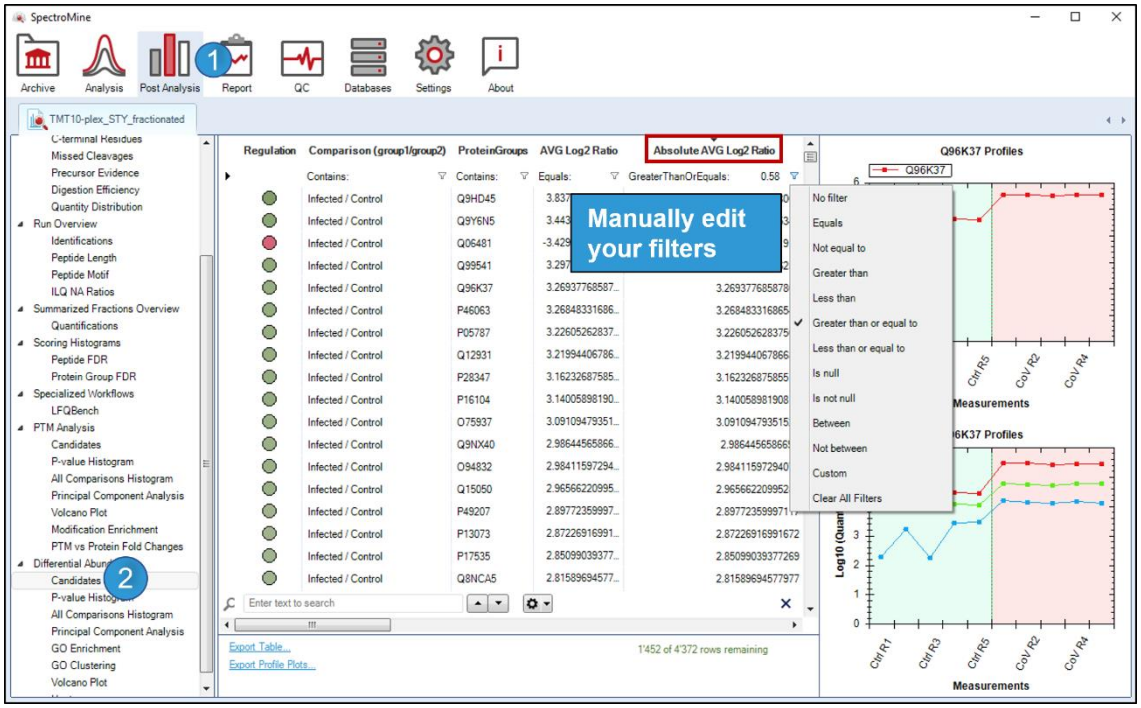

Figure 22. Post Analysis Perspective. Several summaries, tables and plots are available as you navigate through the nodes in the tree. On the figure, a view of the Candidates table. You can modify your candidate set by filtering directly on this table.

3.5.1 Experiment Overview

Here you will find some experiment-wide information that will give you descriptive details about the dataset. Under Overview, you will see a summary with the number of proteins and peptides identified.

To support this overview, you will find a comprehensive set of plots, mainly bar-graphs, explaining your data from several aspects. On each of these plots, you can change some settings by using the right-mouse click option.

3.5.2 Run Overview

Under Run Overview you can find run-specific identifications, peptide length distribution, enriched peptide motifs, and the ILQ NA ratios plot, which displays the percentage of missing values per channel for each run (Figure 23). This plot can be used to verify that channels were correctly loaded across runs. For instance, if you did not use one channel on purpose, this is a great plot to verify that your experiment worked as intended.

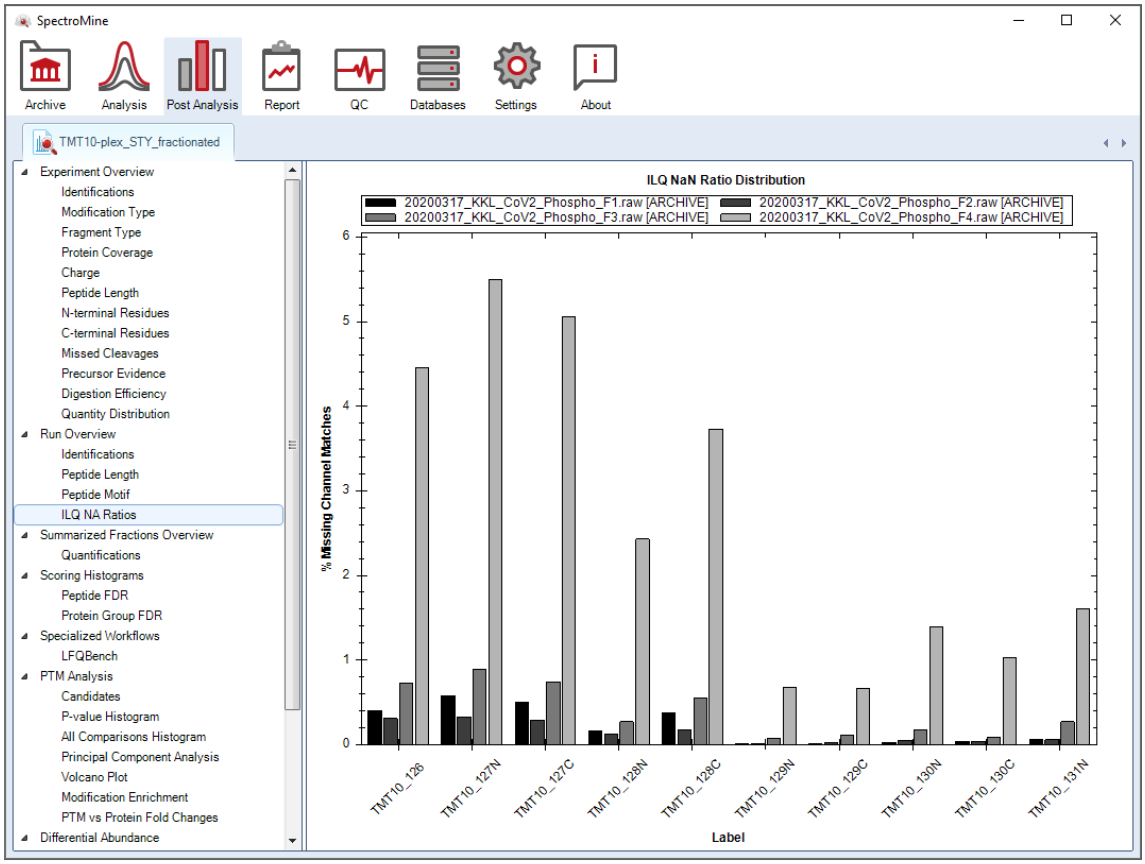

Figure 23. The ILQ NA Ratios plot displays the percentage of missing values per channel for each run.

3.5.3 Summarized Fractions Overview

Under the Summarized Fractions Overview, you can find the sparse and complete quantifications in your experiment. The Protein Groups Quantified in All Channels and the Protein Groups Quantified in at Least One Channel are equal when the ILQ missing value imputation is enabled in the settings.

3.5.4 Scoring Histograms

Under this node, you will find plots related to the behavior of the target and the decoy distributions both at peptide and protein level. This behavior defines the FDR.

3.5.5 Specialized Workflows

Under specialized workflows, a plot dedicated to LFQ benchmark studies is available (Figure 24). It enables visualization of ratios of the mixed proteomes based on experimental data in the function of protein abundance. The plot enables as well to examine how the experimental ratios of proteome diverge from the expected ratios.

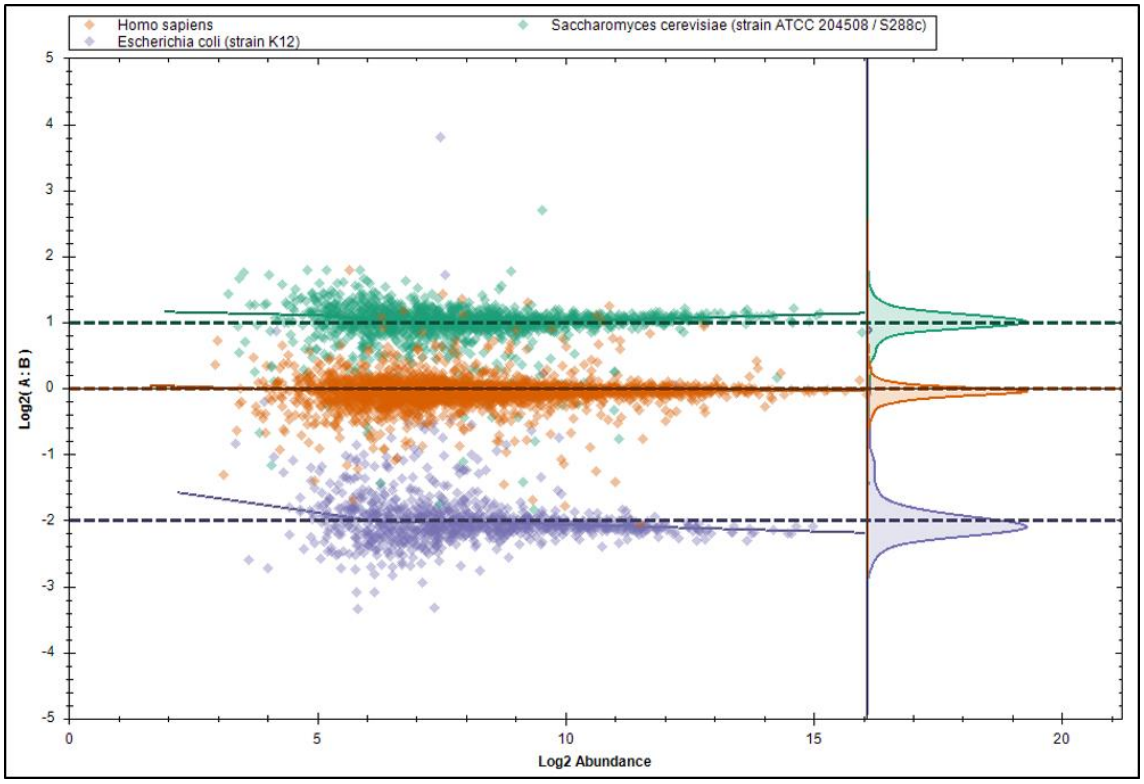

Figure 24. The LFQBench plot for mixed proteomes of three organisms. The Y-axis (log2Ratio) shows the ratios of the mixed proteomes based on experimental data. The expected ratios (dashed lines) can be set by right-clicking on the plot. The distributions on the right side of the plot show how the experimental diverge from the expected ratios. The X-axis represents protein abundance (log2-transformed).

3.5.6 PTM Analysis

A node entirely dedicated to PTM analysis is available when the PTM Workflow is enabled in the Experiment Settings (see section 3.9.2). SpectroMine performs PTM differential abundance analysis on a modification site level.

The results of the PTM differential abundance analysis are reported in a dedicated 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 (Figure 27) and PTM vs Protein Fold Changes (Figure 28).

3.5.6.1 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 (Figure 25). The identification key of the specific modification site object, for which differential abundance analysis result is reported, can be found in the column «Group». In example, O60716_Y291_M3. Where O60716 is the protein identifier, Y291 is the PTM site, and M3 is the multiplicity of 3 for a peptide that carries three modifications.

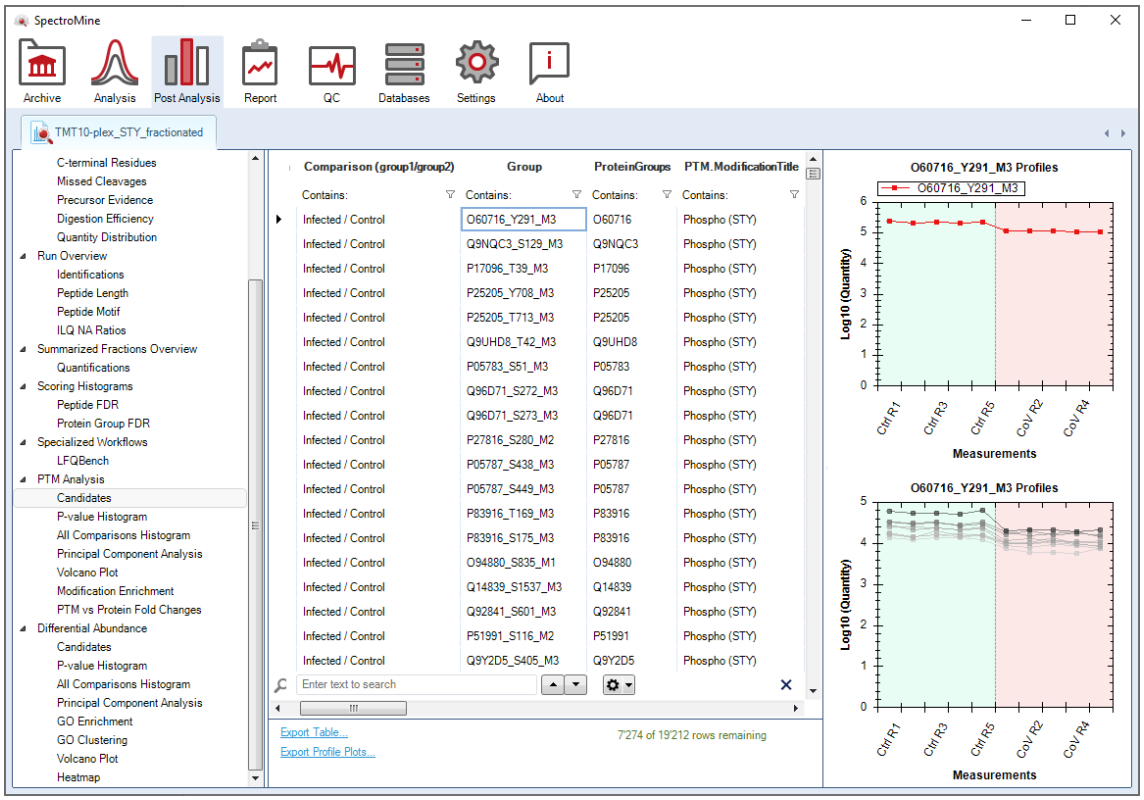

Figure 25. The PTMs Candidates table reports differentialy abundant modification sites. The Group column contains the identification key of the differentially abundant modification site.

The PTM analysis candidates table can be viewed, modified and exported in a similar way as the candidates of the differential abundance analysis at the protein level (Figure 22 and section 3.5.8.1). The candidates of the PTM differential analysis will be visualized in the corresonding volcano plot (Figure 26).

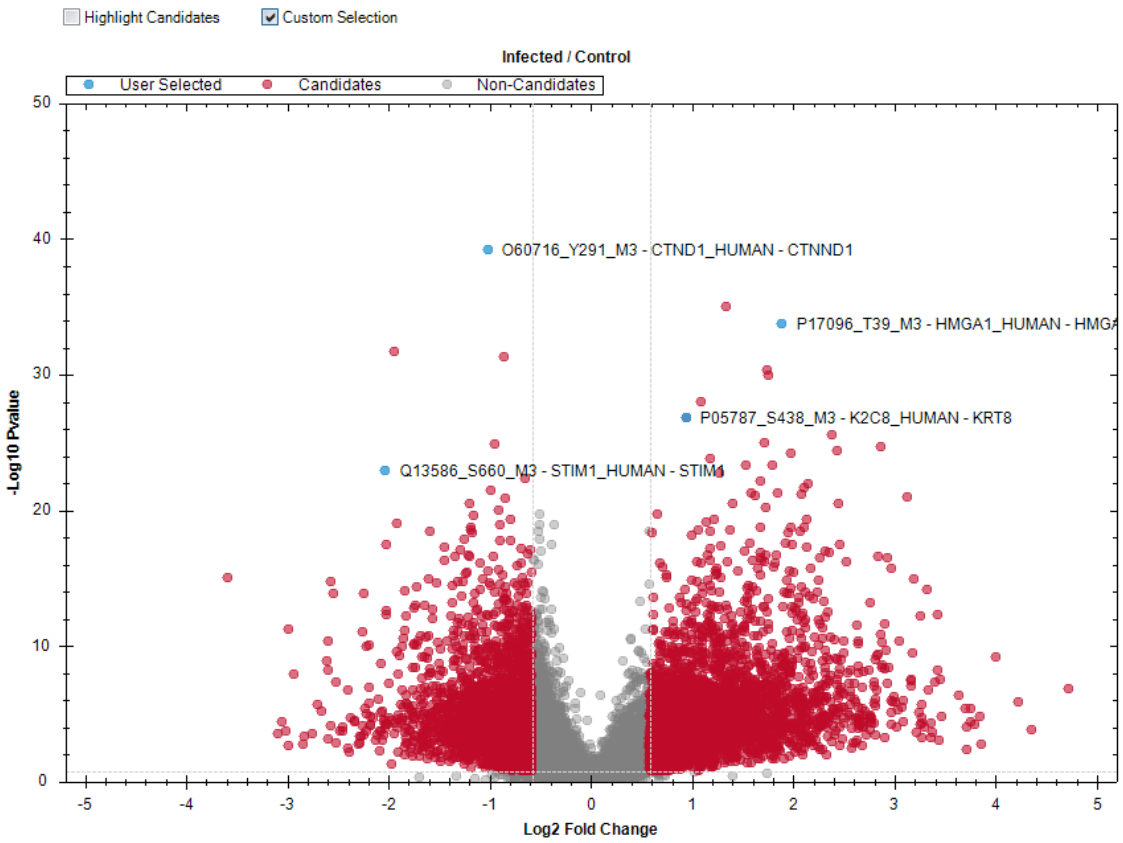

Figure 26. The Volcano Plot of the PTM analysis shows -Log10 P value plotted against Log 2 Fold change of differential abundance on the modified site level. By default, the filters are set to ≥ 1.5 absolute fold change and ≤ 0.05 q-value. Custom Selection can be chosen to highlight only user selected ones.

3.5.6.2 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 inFigure 27, 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.

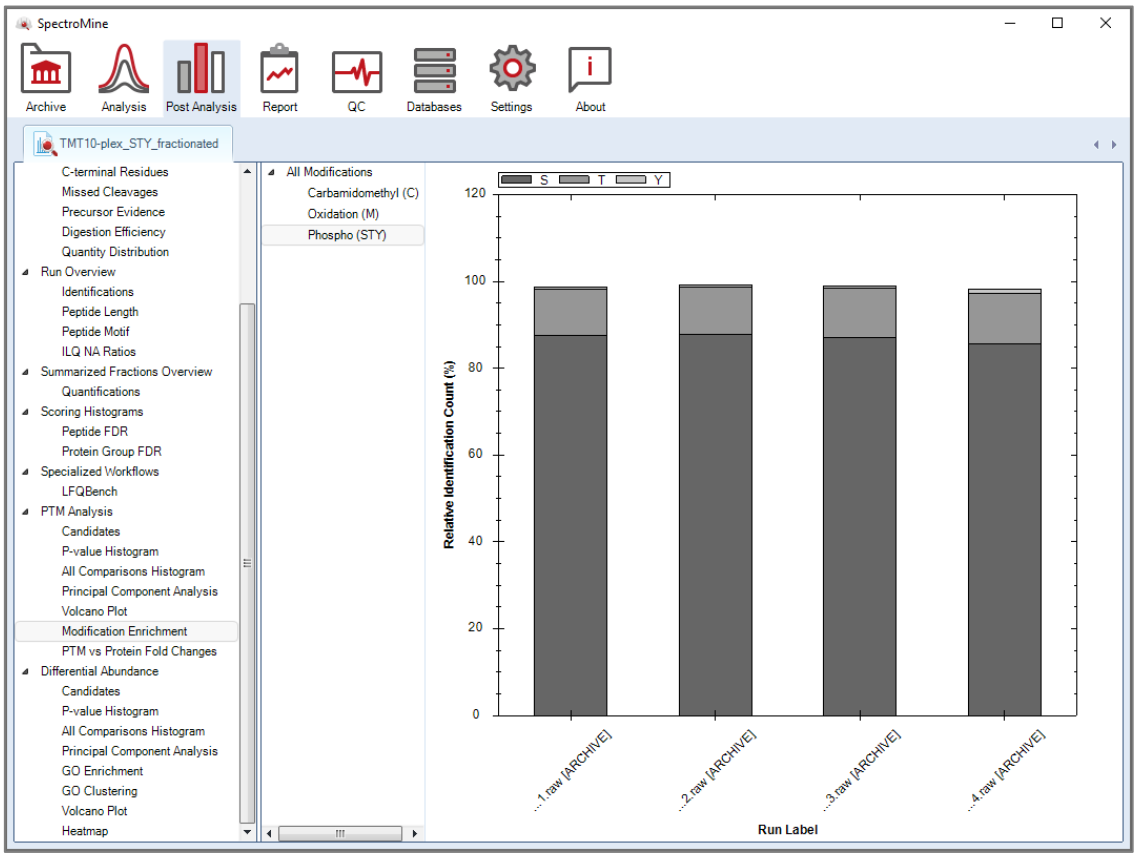

Figure 27. Modification Enrichment plot shows the percentage of all the precursors carrying given modifications in each of the experimental runs.

3.5.6.3 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 identify changes on the PTM site level that are independent of the protein group abundance changes.

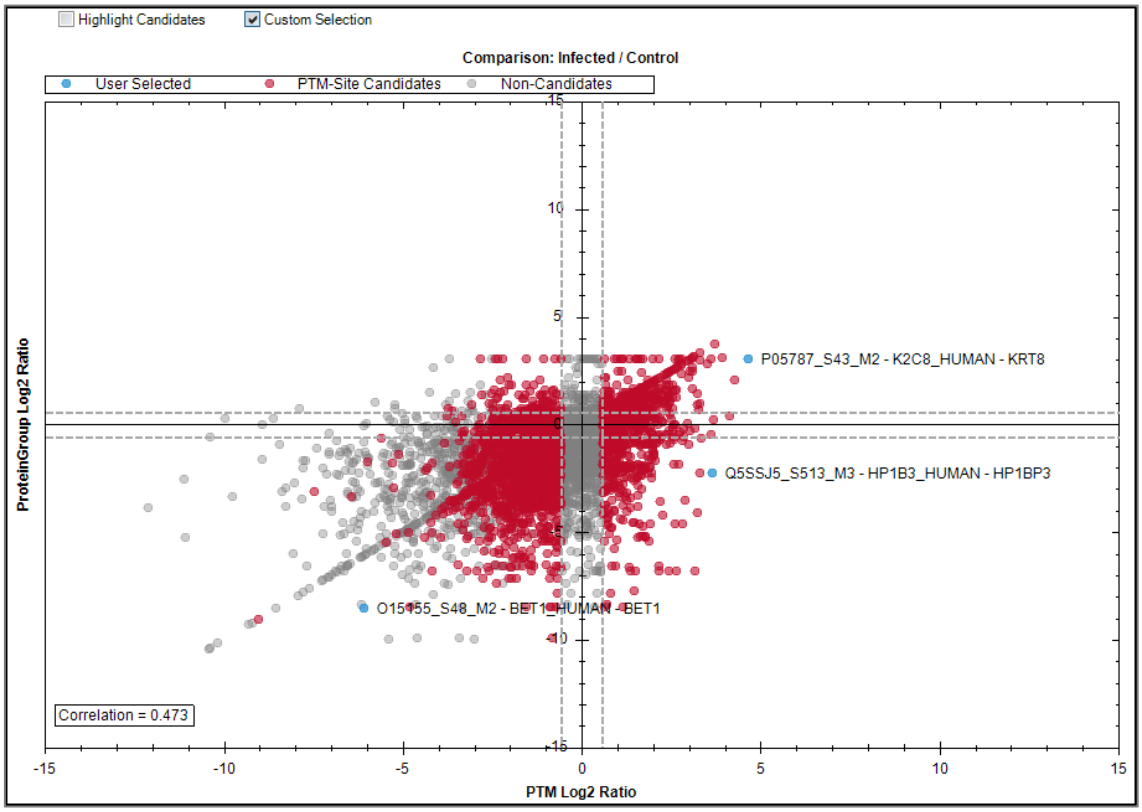

Figure 28. PTM vs Protein Fold Changes plot shows log2 ratio of the protein groups plotted against Log2 ratios of the PTM sites.

3.5.7 Labeling Efficiency

You can visualize chemical labelling efficiency (Figure 34). This plot is available only after choosing labeling efficiency workflow for the analysis of the labeled samples.

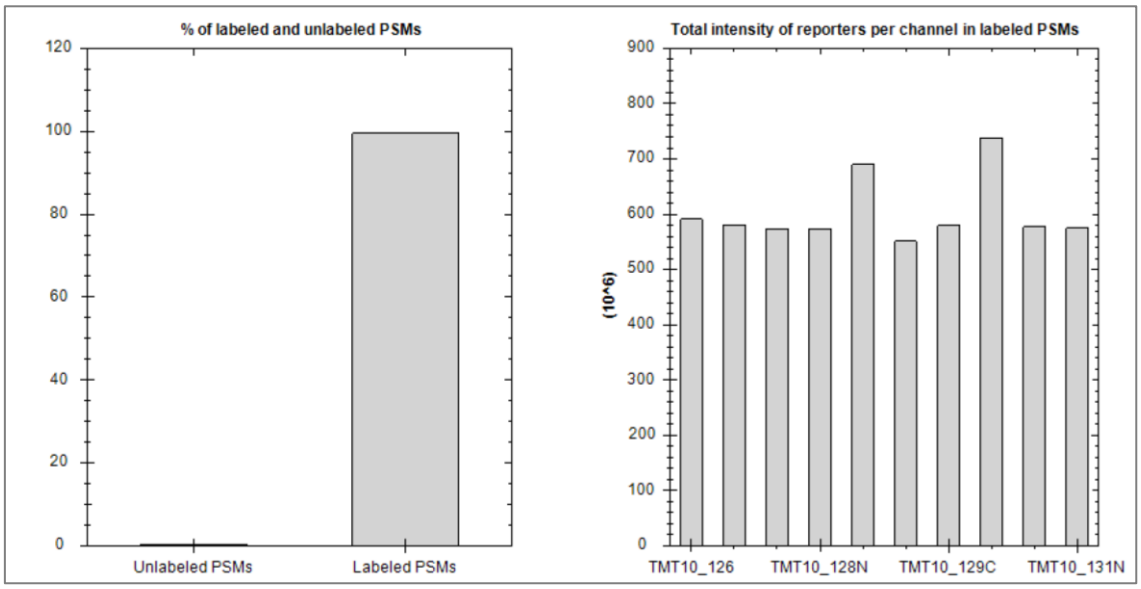

Figure 29 Labeling Efficiency Check results. The left site plot enables visualization of percentage of labeled and unlabeled PSMs. The right site plot shows total intensity of reporter ions per channel in labeled PSMs. Both plots are constructed for each experimental run separately.

3.5.8 Differential Abundance

3.5.8.1 Candidates table

The results from 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 annotated by pair wise comparison (Figure 22):

- The direction and the extent of the 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 log transformed ratios of averaged replicates (AVG Log2 Ratio).

- The table is, by default, filtered by a multiple testing corrected Qvalue 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 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 Pvalue 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 it and choosing your preferred options. You can also export them.

3.5.8.2 Principal Component Analysis

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

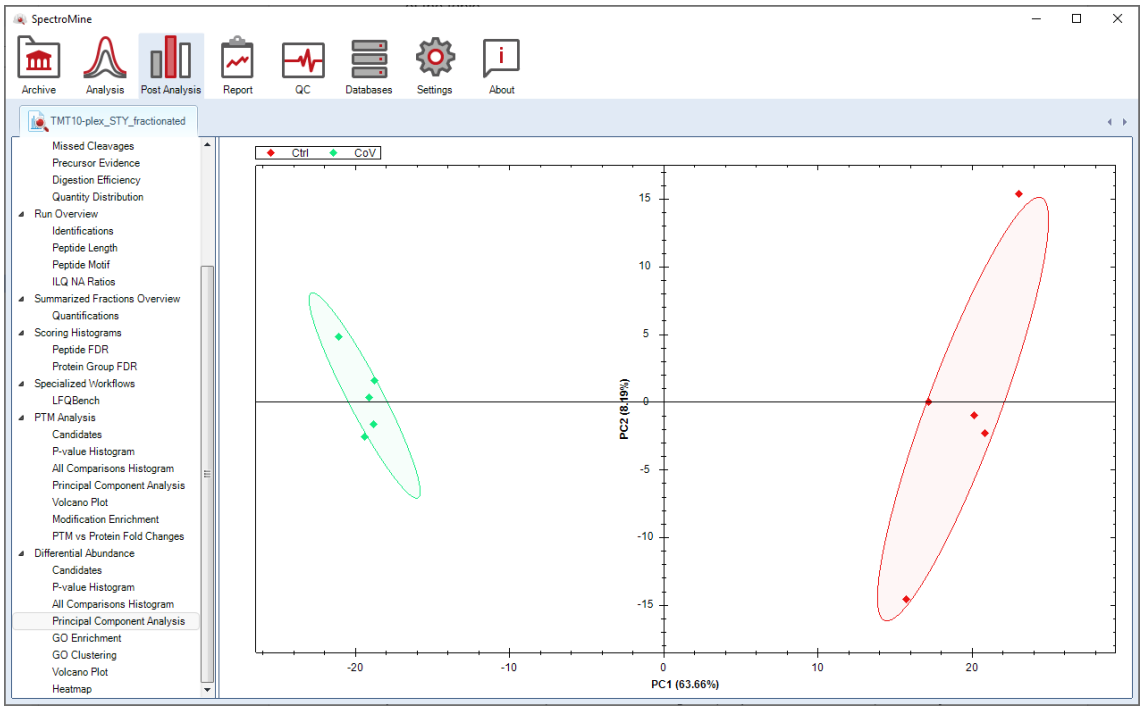

Figure 30 Principal Component Analysis plot. By right clicking on the plot you can perform several actions, such as Save Image As, export the data matrix, or modify the scaling.

3.5.8.3 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, SpectroMine will perform a GO term enrichment test.

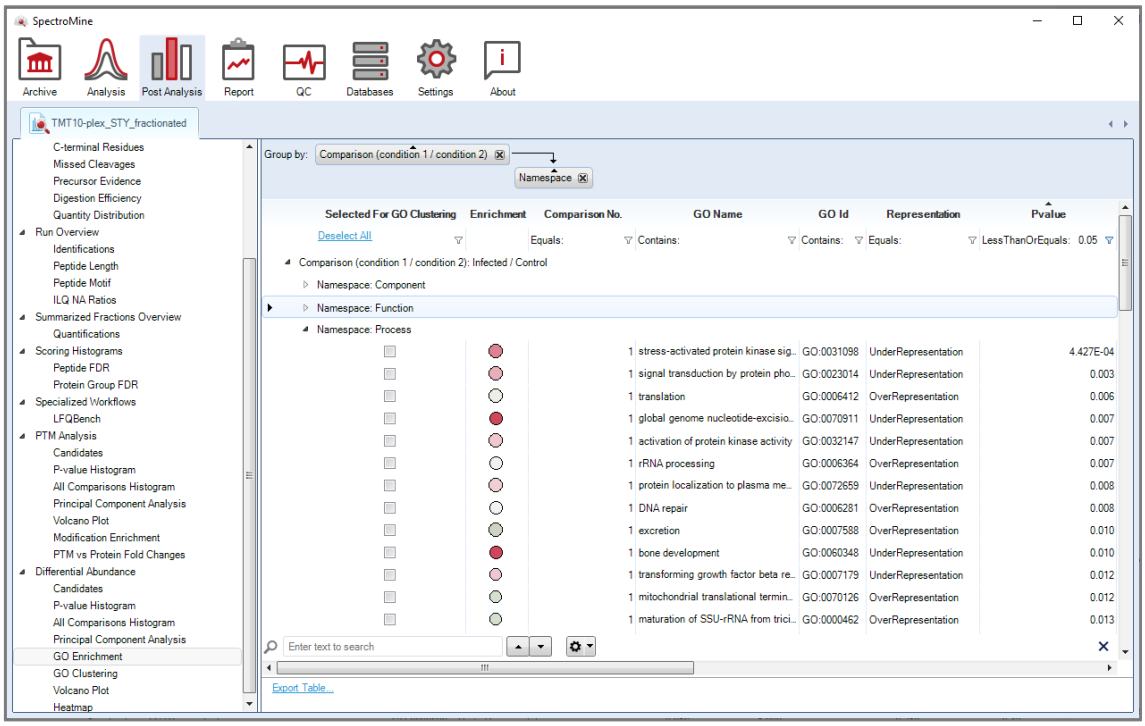

Figure 31. GO term enrichment result. Similar to the Candidates table, you can apply filters. You can also group the results according to any of the column headers by dragging them to the "Group by" field. If you want to do GO clustering on manually selected terms, use the first column of this table.

SpectroMine 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 SpectroMine 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.

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 Pvalue. SpectroMine will perform two multiple testing correction methods to this test: Bonferroni (Dunn 1961) and Benjamini-Hochberg (Benjamini and Hochberg 1995), for which the corresponding corrected Pvalues 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 term enrichment table can be exported by clicking "Export Table…" below the table panel.

3.5.8.4 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 http://revigo.irb.hr/ algorithm (Supek *et al.* 2011). The semantic similarity of two terms is calculated based on their position/relation in the Gene Ontology graph.

SpectroMine 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 pair-wise comparison you want 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 (like 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.

3.5.8.5 Differential Abundance Plots

Under the Differential Abundance node, several plots related to the significance test are also generated. The most relevant are the Heatmap and the Volcano Plot:

1. The Volcano Plot shows the results of the differential abundance test by plotting the peptide or protein fold change against the significance level. The candidates will appear in red on the plot (Figure 32).
2. 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 (**Error! Reference source not found.**).

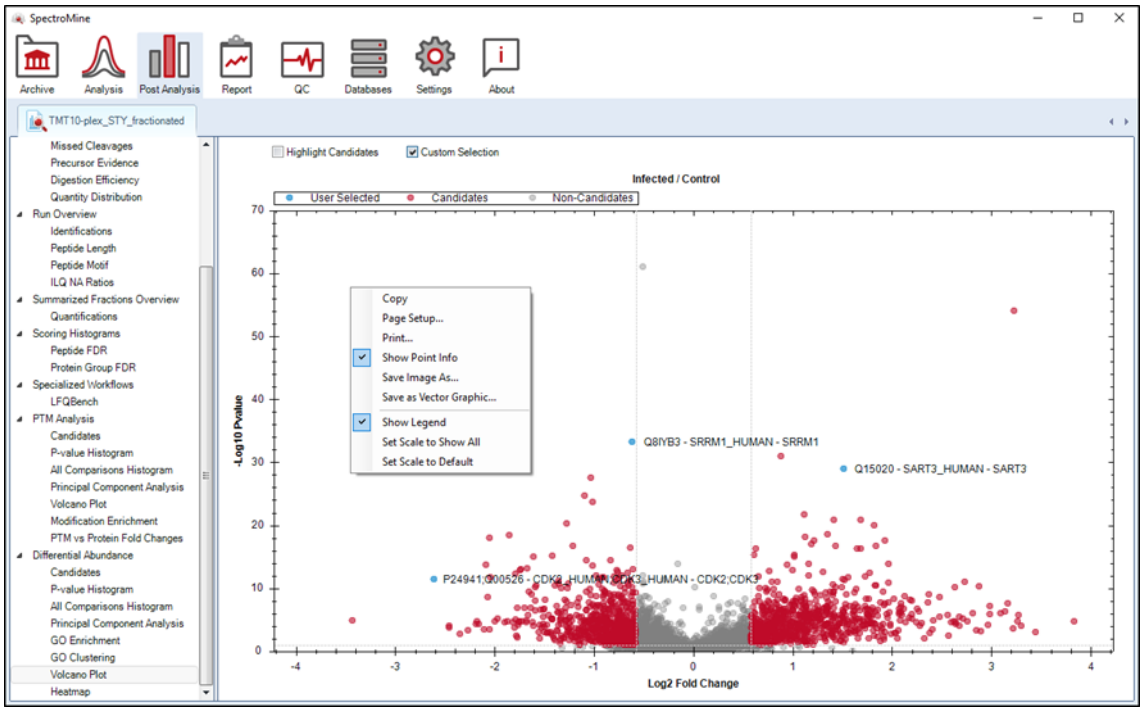

Figure 32 The Volcano Plot shows the candidates in red. This plot is updated when you modify the Candidates table. By right clicking, you can choose several actions, like deactivate the legend or annotate the candidates.

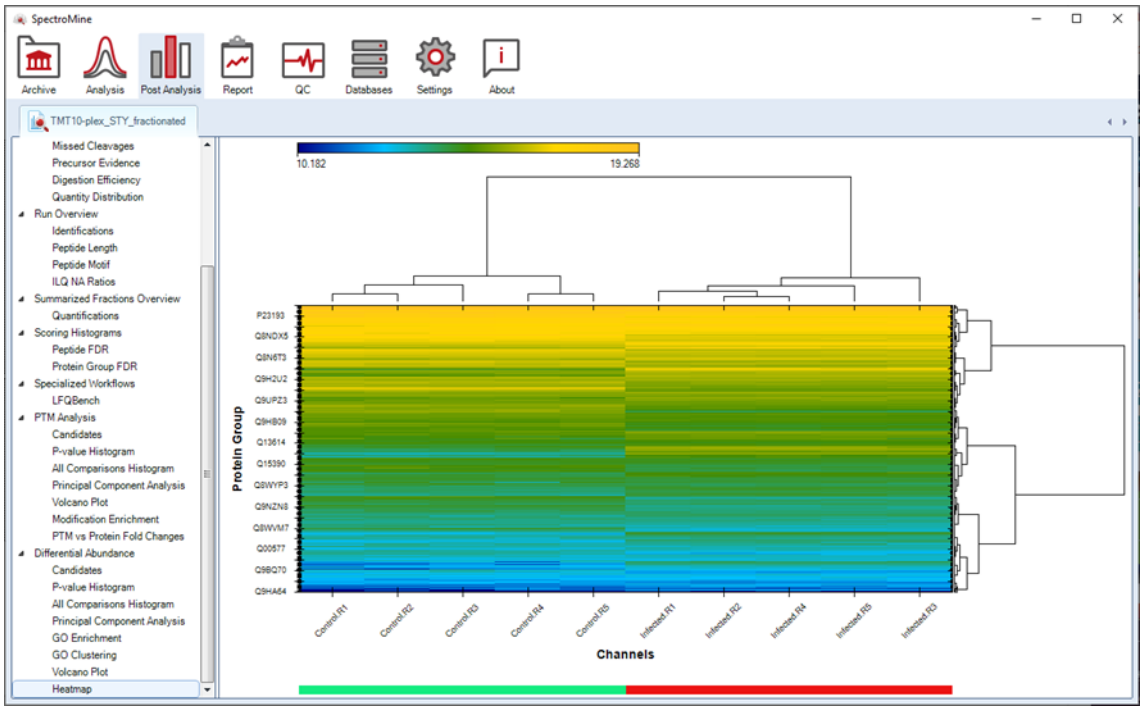

Figure 33 Heatmap with clustering in both rows and columns. The heatmap is built using the set of confidently identified datapoints. By right clicking on the plot, you can perform a number of actions, such as Save Image As, export the data matrix, or modify the scaling.

3.6 Report Perspective

SpectroMine™ has a very powerful reporting tool. In the Report Perspective, you can design and customize your report to contain any information you may need about the analysis. In general, SpectroMine reports are devided into three types that contain information specific to the type of experiment:

- Run Report is available for any experiment analyzed in SpectroMine and it contains qualitative and quantitative information for precursors, peptides and protein groups from each of the experimental runs separately.
- PTM site Report is available when PTM workflow was selected during the analysis. PTM site report contains qualitative an quantitative information at PTM site level. If the experiment contains multiple fractions per sample, the PTM site information will be available only at the sample level (after fraction summarization).
- Summarized Fraction Report is available if samples consist of two or more fractions. Using this report, the user will get qalitative and quantitative information about peptide and protein groups at the sample level.

Report schemas of all report types can be saved and reused. You can also change the column names to fit your needs (Figure 34).

3.6.1 Panels in the Report Perspective

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 headers.

   Many headers have a text hover tool directly in the software with an explanation of what they mean. If you don't find the information you are looking for, do not hesitate to contact us via our Help Center.
3. Filters applied to the report.
4. Report preview: this one will only be visible if there is an analysis loaded. It will show 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.

3.6.2 Report Schemas

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

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

3.6.3 Normal Report

Each of the Report types could be preconfigured as a Normal Report (long format) or as a Pivot Report (wide format). In Pivot Report, you will find each reported event in a single row. Figure 34 shows an example of Normal Report for the Run Report. A Normal Report will often 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 34). The Columns are organized by levels, from more general (Experiment) to more specific (Fragment Ion):

>Experiment
>Run
>Protein Group
>Peptide
>Peptide Precursor
>Precursor
>PSM
>Fragment Ion

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 filed at the bottom of the Columns panel where you can type what you are looking for (Figure 34). Finally, to know which information a header contains, hover over it and you will see a text box popping up with a description.

If you are not sure about the meaning of a column, contact us via our Help Center.

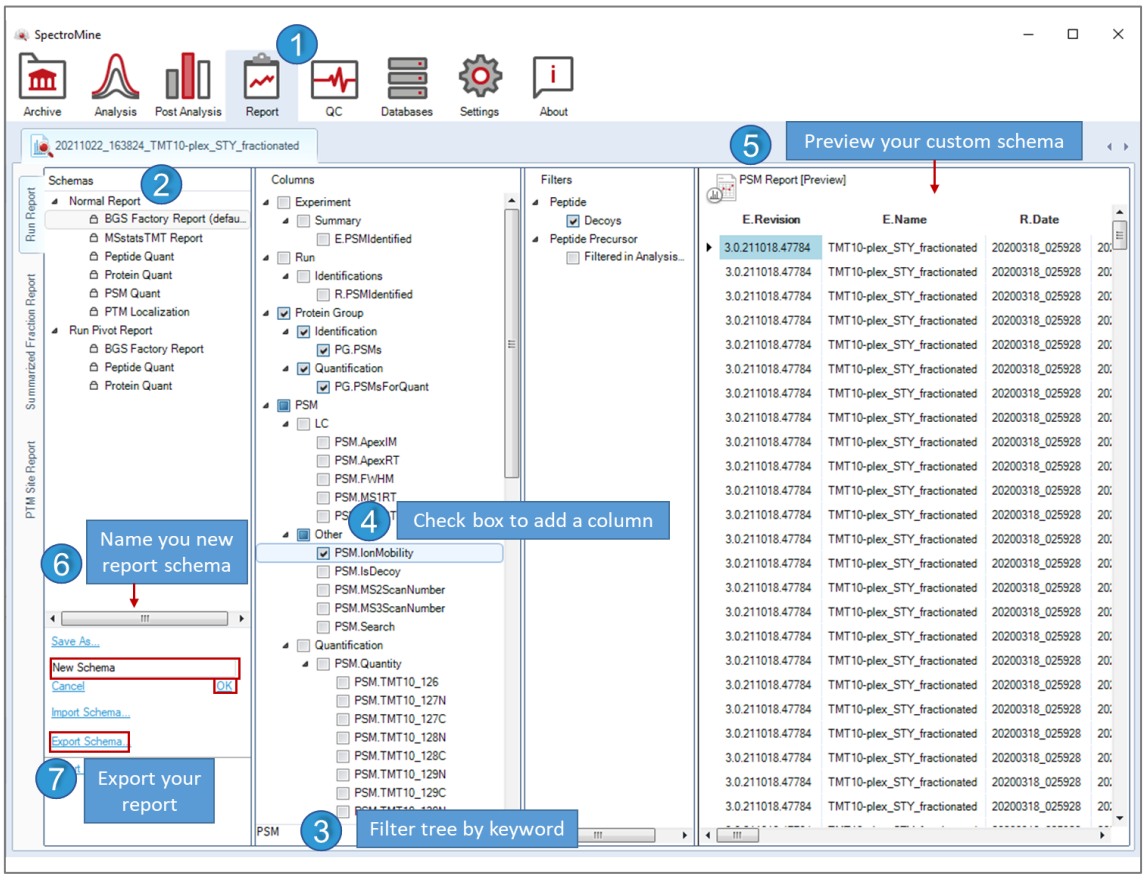

Figure 34. Report Perspective. The figure shows the process of customizing a Normal Report schema and exporting the data. Detailed explanation of the headers can be found by hovering over them.

3.6.4 Run Pivot Report

In a Run Pivot Report (wide format), each run (sample) will be a 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 35). 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.

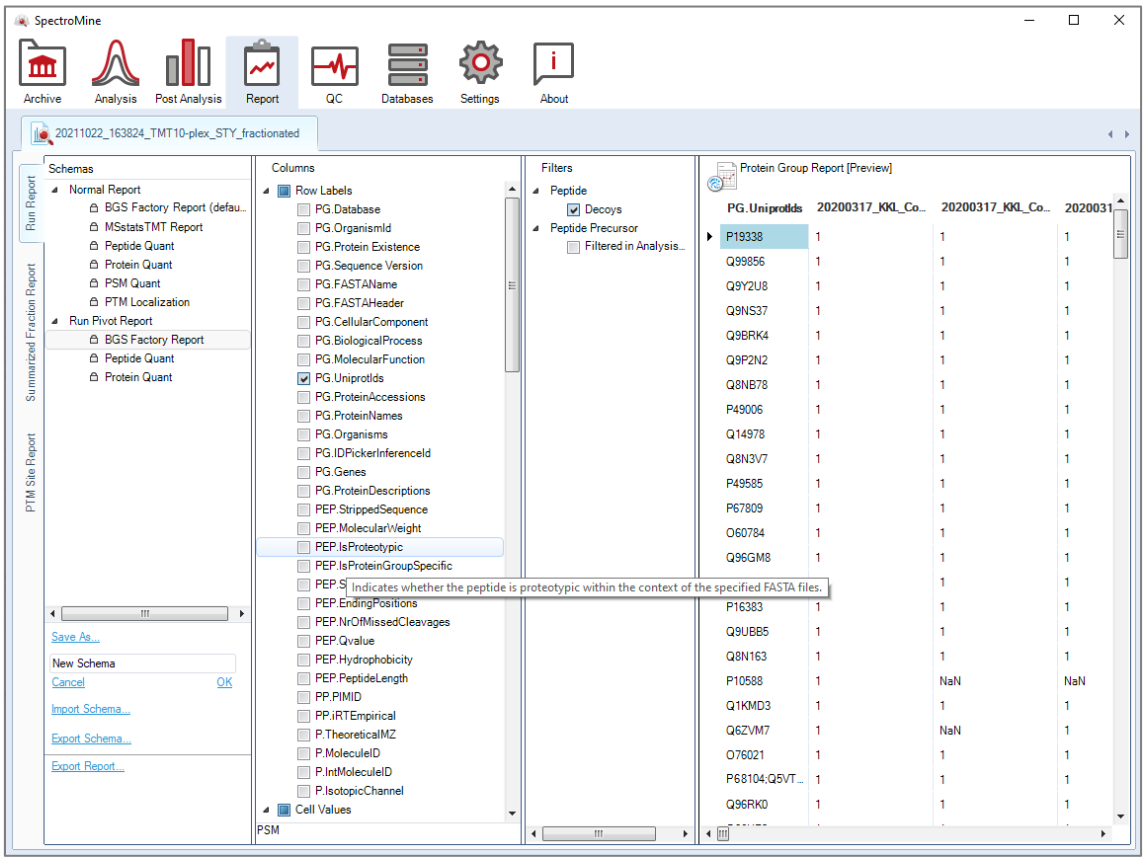

Figure 35. Run Pivot Report. This report is in wide format contains one column per run (sample). Detailed explanation of the headers can be found by hovering over them.

3.6.5 PTM site report

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

>Experiment

>Run (or sample if the experiment contains specified fractions)

>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.

3.6.6 Summarized Fractions Report

If the experiment consist of fractionated samples, then the summarized fractions report will be available. The Columns are organized by levels, similarly like standard report, from more general (Experiment) to more specific (Peptide Group):

>Experiment
  >Sample
    >Protein Group
      >Peptide Group

Among other information, the summarized fractions report contains quantitative and qualitative information on peptide group and protein group levels that was summarized according to fraction summarization settings chosen in the quantification node of the experiment settings.

3.7 QC (Quality Control) Perspective

The quality control perspective of SpectroMine™ allows you to monitor performance indicators across and within experiments (Figure 36). Chromatography, mass spectrometer behavior, and search aspects can be measured and recorded over time. SpectroMine automatically detects various instruments and acquisition modes, 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 testing performed in a specific laboratory.

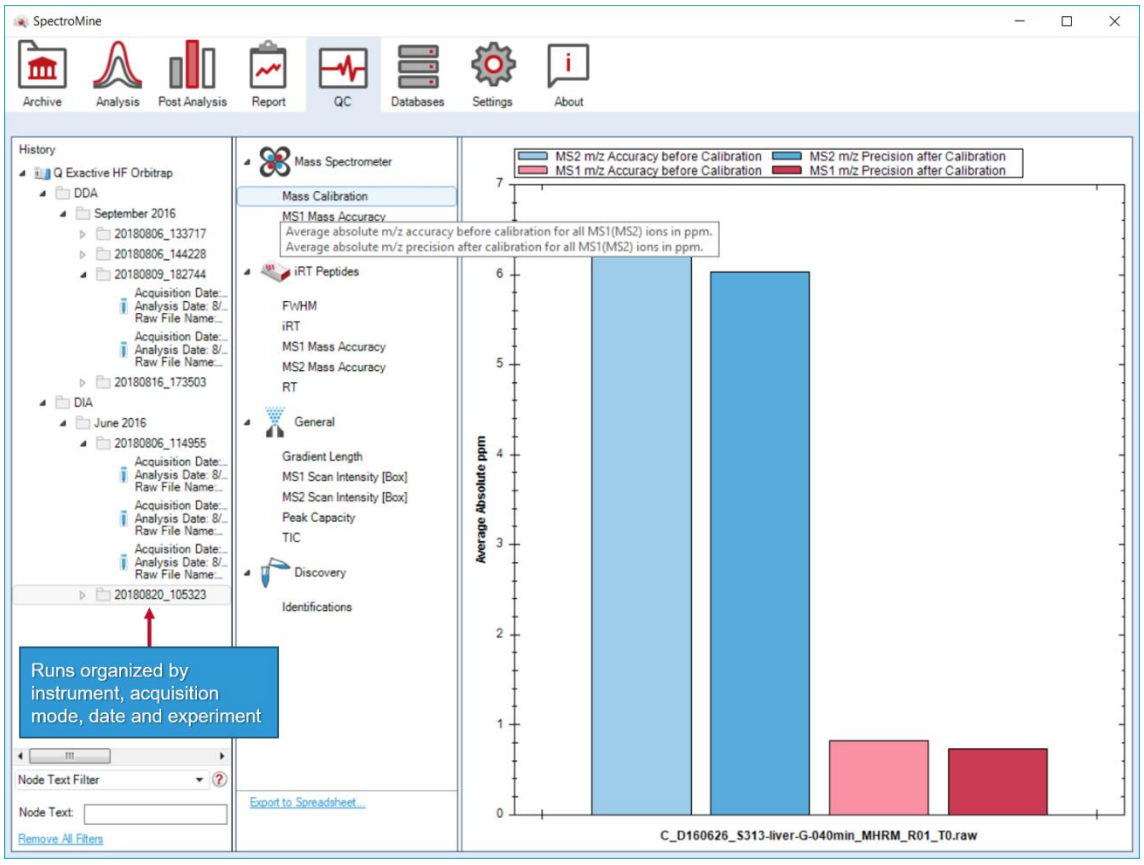

Figure 36. QC Perspective. Runs are saved in the History tree. You can monitor instrument performance thanks to the various plots related to several aspects of the experiment, from LC-MS metrics to identifications.

3.8 Databases Perspective

The Databases Perspective in SpectroMine™ allows you to store and manage information that you will need when setting up analyses. This includes protein databases, gene annotations, peptide modifications, etc. In general, the elements contained in the Databases Perspective will be selectable through the different windows of the set-up wizards.

3.8.1 Protein Databases

This section of the Protein Databases Perspective lets you import and manage your protein databases. SpectroMine uses protein databases (FASTA files) to make searches and to perform protein inference. The imported protein databases contain all the sequences, as well as meta-information extracted from the FASTA protein headers, using the specified parsing rule. SpectroMine already contains the UniProt parsing rule, but you can add a new rule by clicking "New Rule" in the Protein Databases overview or during an importation (Figure 37).

To import a new proteome database from FASTA click "Import…" in the bottom-left corner" (Figure 37). While importing a new protein database from FASTA, SpectroMine will try to find the appropriate parsing rule for this file format from the already specified rules. Should there be 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 for setting up an analysis.

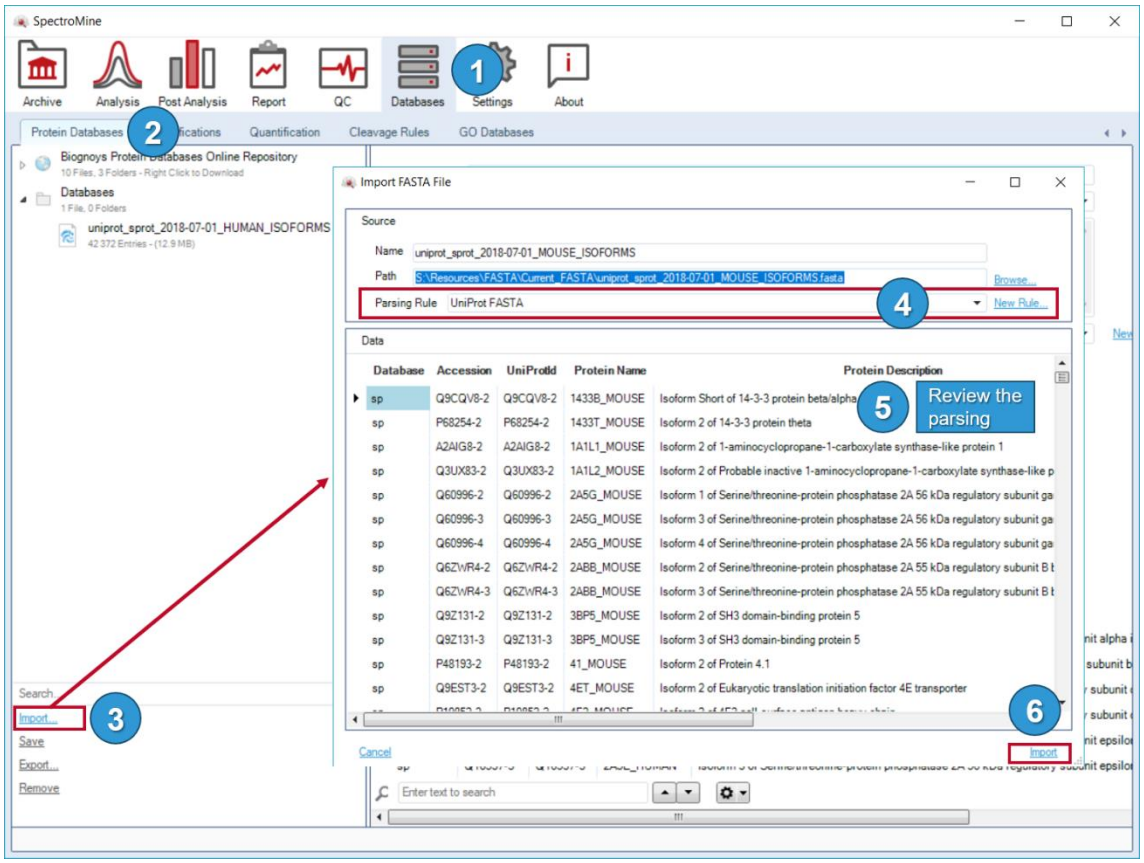

Figure 37. Importing a new FASTA file into SpectroMine. Your new database will appear in the Databases tree and will be available for setting-up analyses.

3.8.2 Modifications

The SpectroMine modifications panel provides a central location to manage modifications. Modifications contained here can be added to the Pulsar Search Settings when setting up an experiment. SpectroMine comes with a database of default modifications. If you use special modifications, add them to the Modifications database using this Perspective. You can also modify an existing modification or add more information to it.

To create a new modification, click "new" in the bottom-left corner, give a name to your modification and click "OK" (Figure 38). 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…".

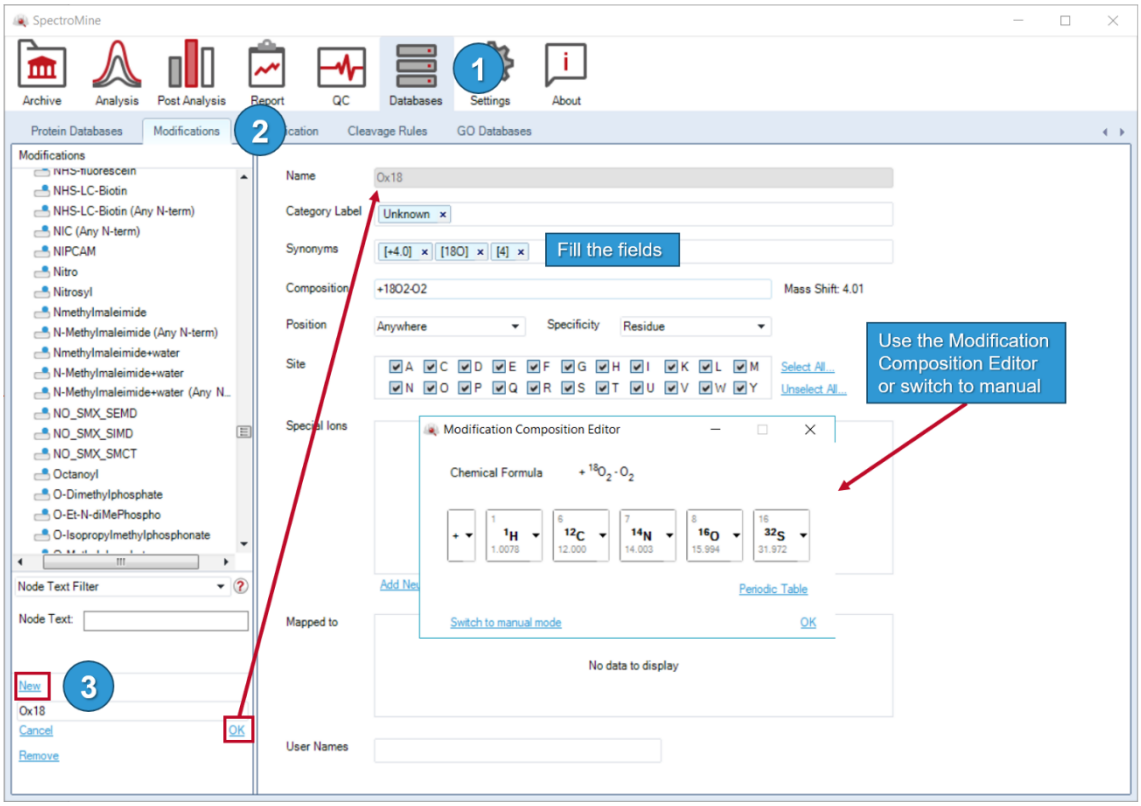

Figure 38. Adding a new modification to the database. Click New and give a name. Specify which sites does it apply to and fill in the composition. Click OK.

3.8.3 Quantification

In the Quantification page of the Databases Perspective, you will find your different isobaric label quantification schemas. This page includes a tool to build you own schema (Figure 39). To do that, click "New" in the bottom-left corner and give a name to your schema. Create your channels by specifying the chemical formula of each reporter ion and its charge. Next, specify to which modification the isobaric label should be associated with.

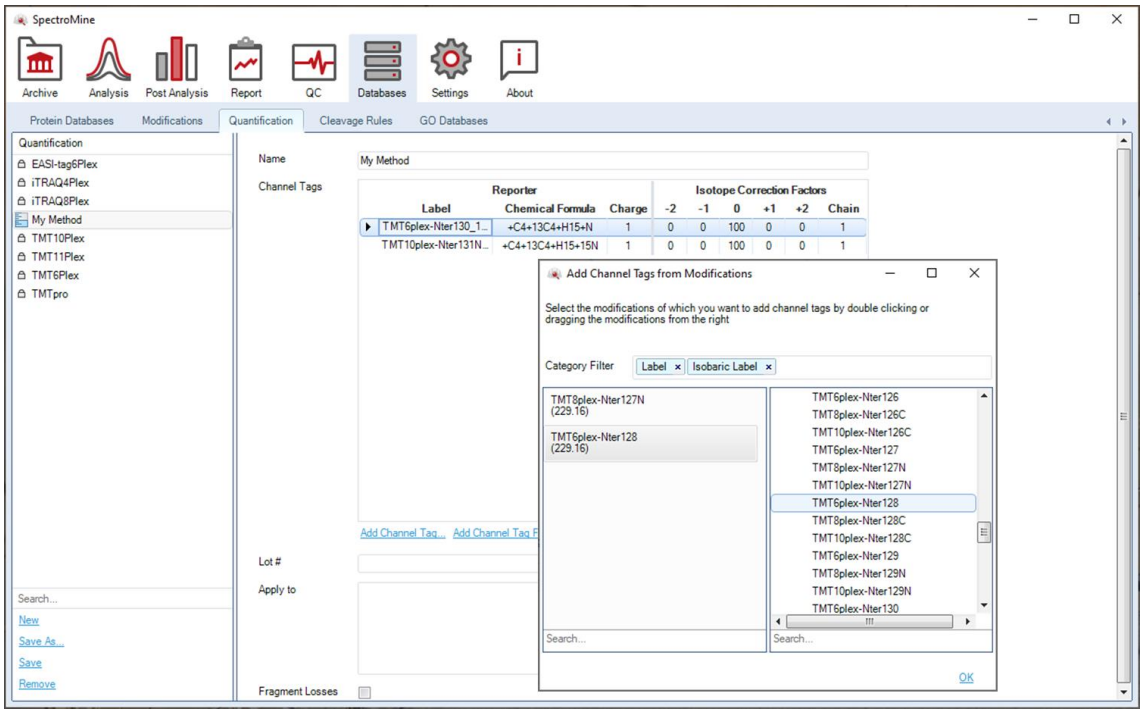

Figure 39. Add a custom ILQ method. Add your channels from modifications or type the formula manually. Specify to which modifications it should apply (see Figure 11, top panel).

If you need help creating your isobaric label quantification schema, contact us via our Help Center.

3.8.4 Cleavage Rules

The tool contained in this page lets you define the rule to *in-silico* digest your proteins from the protein database(s) and create the proper search space. Cleavage rules are applied whenever you do a search in SpectroMine. The most frequent rules are already included in SpectroMine, such as Trypsin, Trypsin/P, and LysC.

To design your own rule, you must select an existing one, modify it and click "Save as…" in the bottom-left corner (Figure 40). The rules are defined by which sites are cleaved by the enzyme. In the Cleavage 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 40). At the bottom, you can see a preview of how a sequence will look like after being cleaved following your digest rule. You can also include a description.

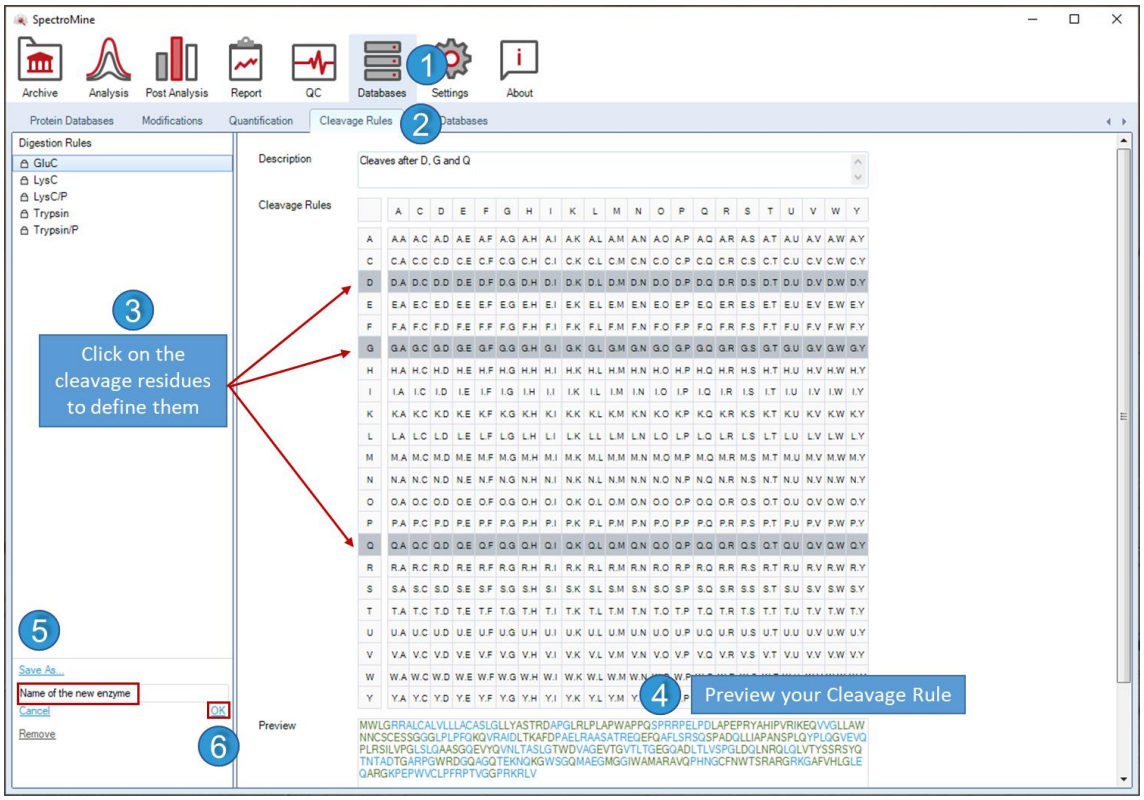

Figure 40. Define a new Cleavage Rule. The Cleavage Rule editor will allow the generation of new cleavages rules in a very friendly manner.

3.8.5 GO Databases

Similar to the Protein Databases, gene ontology (GO) Databases in SpectroMine are used to further annotate your data with biological information. 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.

3.8.5.1 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, SpectroMine supports the *.obo file format from the GO Consortium. The go-basic.obo is already part of the SpectroMine installation. Information from a gene ontology tree can only be used in combination with an organism specific gene annotation file.

3.8.5.2 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, SpectroMine 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 also specify a custom annotation file.

The easiest way to obtain a Gene Ontology annotation is by downloading it from the Biognosys Online repository (Figure 41).

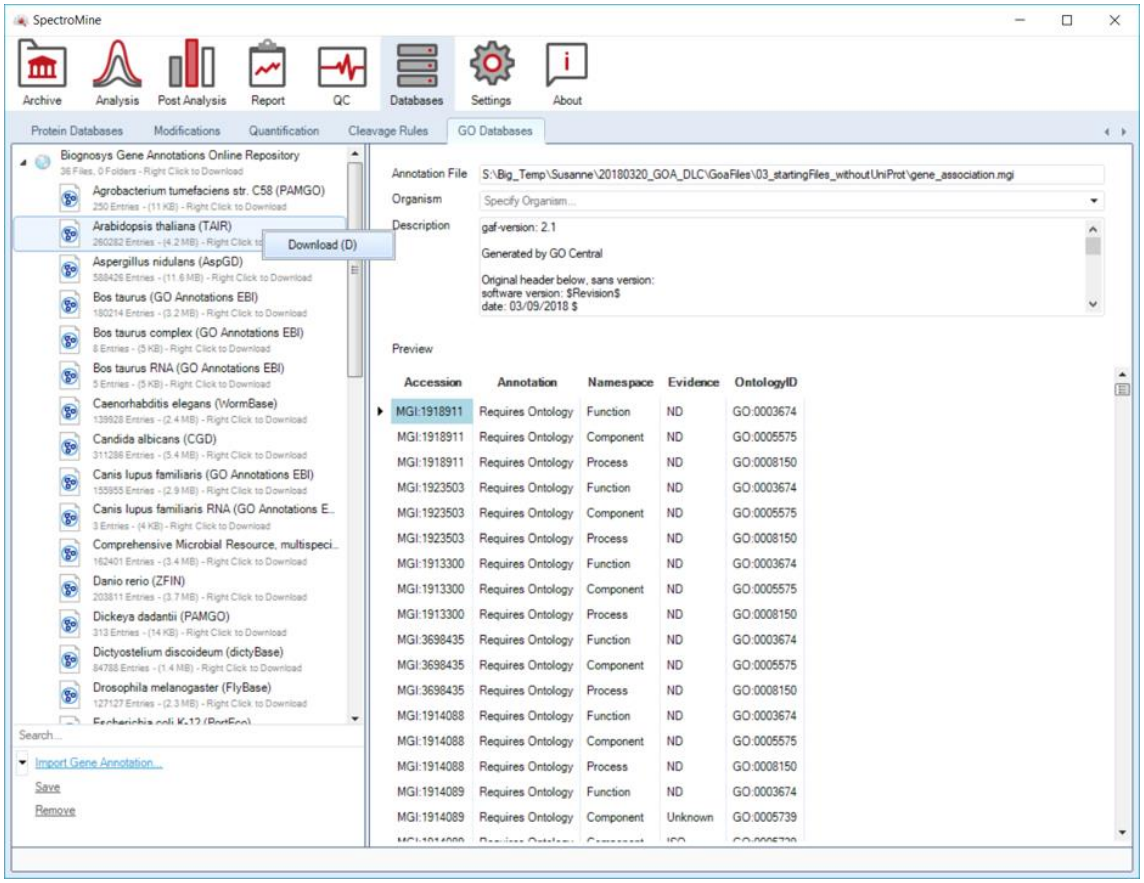

Figure 41. GO page of the Databases Perspective. The easiest way of obtaining a new Gene Ontology annotation is by downloading it from the Biognsosys online repository. Expand that element and right click the node corresponding to your organism to download the annotation.

To import a new gene annotation file into SpectroMine, 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.

3.9 Settings Perspective

The Settings Perspective of SpectroMine™ 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: Search Settings and Experiment Settings (Figure 42). In addition, you can alter global settings of SpectroMine 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 42).

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 42). Your new schema will appear in the tree and it will be available to be chosen during the set-up of your next analysis.

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

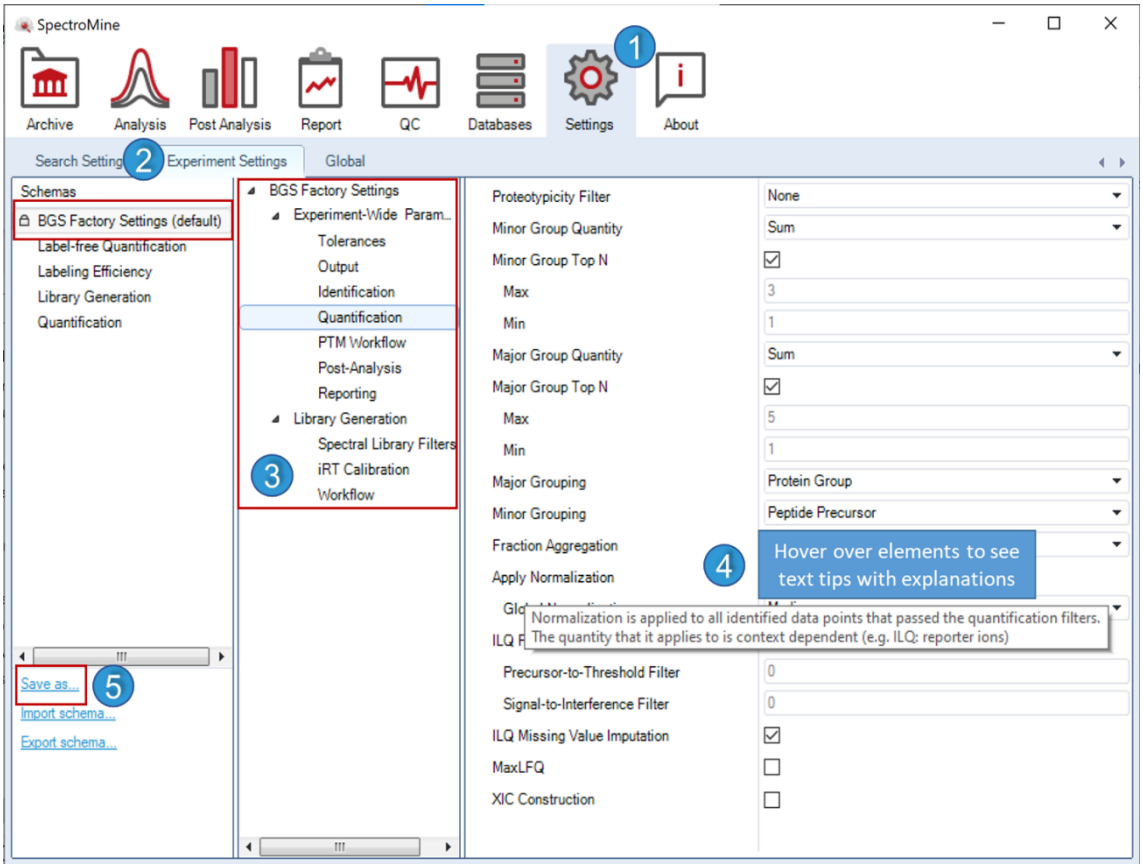

Figure 42. Make a custom schema for your analysis. The new schemas will be available during the subsequent analysis set-ups.

3.9.1 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 Appendix 1. Pulsar Search Settings (section 5.1).

3.9.2 Experiment Settings

The Experiment Settings define the details of how SpectroMine should analyze the data. These settings will specify important metrics, such as which output to generate, FDR cutoffs, quantification setting, filters for your spectral library, among many others. Find details of each setting in Appendix 2. Experiment Settings (section 5.2).

3.9.3 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.

3.9.3.1 General

This section contains settings options that allow you to modify the default behavior of SpectroMine. For more information about these options use the tool-tip hover for each individual entry. If still in doubt, contact us via our Help Center.

3.9.3.2 Directories

Here you can setup the different storage paths for data managed by SpectroMine. Please note that all changes within the "Directories" section will require a restart of SpectroMine to take effect. See some recommendation in section 1.4.

3.9.3.3 Visualization

The plotting section allows you to customize the look of your plots by changing the line strength. You can also define here number of runs for which visualization will be switched to “Many Runs” mode.

3.9.3.4 Reporting

Specify a location to save the reports generated using the "Export All…" function in the Experiment Tab. Also, define the file naming strategy for those reports.

3.10 SpectroMine™ Command Line Mode

In addition to the visual pipeline mode, SpectroMine can also be executed from command line. To run SpectroMine in command line mode you simply call the Spectromine.exe file using the following parameters.

3.10.1 General Arguments

Argument	Explanation
-h	Shows an overview of all command line arguments.
-n	[Optional] Name of the experiment.
-r	[Multiple] Path to run 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	[Multiple] Path to run folder Adds all recognized raw files 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.
-regex	[Optional] Regular expression filter Applies a regular expression as filter to the -d command.
-ext	[Optional, Multiple] Run file extension filter Only keeps the run files from the -d command that match the extensions given.
-s	[Optional, Default] Experiment Settings schema Specifies the Experiment Settings 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 SpectroMine Experiment Settings schema repository (i.e., you should see it in the GUI).
-o	[Default] Output directory Specifies the output directory in which the reports will be generated. The reports will be located in an experiment-specific subfolder within this output directory. The default location is specified in Global Settings -> Reporting -> Batch Output Destination.
-a	[Default] Path to final search archive 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	[Default] Path to final library 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.
-srs	[Multiple, Optional] Path to report schema Specifies the path of the report schema to be generated for this experiment. The report schemas can be personalized in the SpectroMine user interface (Report Perspective).
-con	[Optional] Condition Setup Specifies the condition setup to be used for the post processing (such as regulation analysis). This condition setup file is best generated from the SpectroMine user interface and then exported for use in the command line mode.
-q	[Optional] Quantification Method In case you do not specify a condition setup but are interested in ILQ quantification values, you can assign a quantification method

	using -q. The quantification method can be added using the name that is visible in the SpectroMine user interface
-go	[Multiple, Optional] Path to Gene Annotation file
-command	Path to a file that contains all the arguments Instead of specifying all arguments in the command line, it is also possible to put them in a .txt file and specify the location of the text file using -command.

3.10.2 Arguments specific for FASTA-based search

Argument	Explanation
-fasta	[Multiple] Path to FASTA file Specifies the path to a *.bgsfasta (Managed FASTA, namely FASTA file containing parsing rule i.e. a set of instructions that inform SpectroMine on how to read the column headers) file to be used as the DDA search space. The default location for Managed FASTA files is configured in the Settings Perspective > Global > Directories.
-fastar	[Multiple, Optional] Path to FASTA file for the last run/directory Specifies the path to a *.bgsfasta (Managed FASTA, namely FASTA file containing parsing rule i.e. a set of instructions that inform SpectroMine on how to read the column headers) file to be used as the DDA search space, but only for the last run/directory that was added to the experiment. The default location for Managed FASTA files is configured in the Settings Perspective > Global > Directories.
-rs	[Default] Search Settings schema Specifies the Search Settings 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 Search Settings schema repository (i.e., you should see it in the GUI).
-rsr	[Default, Optional] Search Settings schema for the last run/directory Specifies the Search Settings to be used for the last run/directory that was added to the experiment. 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 Search Settings schema repository (i.e., you should see it in the GUI).
-sa	[Multiple, Optional] Path to Search Archive Adds a specific Search Archive (*.psar file) to this experiment. The default location for Search Archives is configured in the Settings Perspective > Global > Directories.
-sad	[Multiple, Optional] Path to Search Archive directory Adds all search archives (*.psar files) within a specified directory to this experiment.

3.10.3 Arguments specific for library-based search

Argument	Explanation
-l	[Multiple] Library path In case of a library-based search, the libraries can be specified using this command. Either FASTA or libraries should be specified.

3.10.4 Example

```
Spectromine.exe -d "C:\data\My Experiment" -fasta "C:\data\My Experiment\my_organism.bgsfasta"
```

```
-rs " C:\data\my_search_settings" -o "C:\data\My Experiment\Results" -n "My Experiment"
```

If you encounter problems trying to run SpectroMine from command line, please contact us via our Help Center.

4 References

Amodei D, Egertson J, MacLean BX, Johnson R, Merrihew GE, Keller A, Marsh D, Vitek O, Mallick P, and MacCoss MJ (2019) Improving Precursor Selectivity in Data-Independent Acquisition Using Overlapping Windows. *J Am Soc Mass Spectrom* 30:669–684.

Bekker-Jensen DB, Bernhardt OM, Hogrebe A, Martinez-Val A, Verbeke L, Gandhi T, Kelstrup CD, Reiter L, and Olsen J V (2020) Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat Commun* 11:787.

Benjamini Y, and Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc* 57:289–300.

Bruderer R, Bernhardt OM, Gandhi T, and Reiter L (2016) High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. *Proteomics* 16.

Dunn OJ (1961) Multiple Comparisons among Means. *J Am Stat Assoc* 56:52–64.

Escher C, Reiter L, MacLean B, Ossola R, Herzog F, Chilton J, MacCoss MJ, and Rinner O (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* 12:1111–1121.

Mi H, Muruganujan A, Casagrande JT, and Thomas PD (2013) Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc* 8:1551–1566.

Supek F, Bošnjak M, Škunca N, and Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6:e21800.

Tsou CC, Avtonomov D, Larsen B, Tucholska M, Choi H, Gingras AC, and Nesvizhskii AI (2015) DIA-Umpire: Comprehensive computational framework for data-independent acquisition proteomics. *Nat Methods* 12.

Virreira Winter S, Meier F, Wichmann C, Cox J, Mann M, and Meissner F (2018) EASI-tag enables accurate multiplexed and interference-free MS2-based proteome quantification. *Nat Methods* 15:527–530.

Zhang B, Chambers MC, and Tabb DL (2007) Proteomic parsimony through bipartite graph analysis improves accuracy and transparency. *J Proteome Res* 6:3549–3557.

5 Appendixes

5.1 Appendix 1. Pulsar Search Settings

Configure the conditions for SpectroMine™ to perform the search:

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: only one 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 removing the N-terminal M (when there is one)

Labeling

You can specify up to three different channels with assigned labels (modifications)

Modifications

Do Modification Localization	Specify if you want to have modification localization probability for the variable modification
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

MS1

You can specify maximum number of MS1 features searched for 1 MS1 scan.

MS2

You can specify which ions should be searched (ions generated in HCD, CID and EthcD fragmentationAlso, you can choose here option to perform demultiplexing which allows the processing of alternating shifted MS2 windows as presented by D. Amodei and colleagues (Amodei *et al.* 2019)

MS3

When checked, quantification will be done on MS3 if there are MS3 scans. Otherwise, MS2 will be used for quantification.

Speed-Up

Max Variable Modifications	Limit variable modifications in calibration search to the chosen number.
Calibration Search Always Specific	Generate the search space for the calibration search assuming specific enzyme digestion independently of which specificity was set for the main search
MS2 Index	Specifies if MS2 Index is applied. 1. Automatic – MS2 index is used only for DDA runs and in the following cases: - when search space is heavily modified (>2 variable modifications or phosphorylation - if the MS2 index already exists - if there are at least 3 runs in the experiment. 2. MS2 Index could be manually be switched on or off

5.2 Appendix 2. Experiment Settings

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

5.2.1 Experiment-Wide Parameters

Tolerances

SpectroMine™ will, by default, calculate the ideal mass tolerances to generate the library. SpectroMine 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. SpectroMine will do this under default settings (Dynamic).

However, SpectroMine 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 SpectroMine 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

Output

Specify which elements should be generated as a result of the experiment: Search Archive and spectral library

Identification

PSM FDR	Set False Discovery Rate on the peptide spectrum match level
Peptide FDR	Set False Discovery Rate on the peptide level
Protein Group FDR	Set False Discovery Rate on the protein group level
Exclude Single Hit Proteins	Select if you want to exclude proteins that were identified by only one stripped sequence across entire experiment

Quantification

Proteotypicity Filter	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
Minor (Peptide) Group Quantity	Specify how the peptides should be quantified from lower level elements. Choose among sum, median, mean and geometric mean
Minor (Peptide) Group Top N	Use a specific range of the best lower level elements to calculate the minor group quantities
Major (Protein) Group Quantity	Specify how you want peptides to be used to calculate protein quantities. Choose among sum, median, mean and geometric mean
Major (Protein) Group Top N	Use a specific range of the best peptides to calculate proteins
Major (Protein) Grouping	Specify what should be considered as a protein (Protein Group).
Minor (Peptide) Grouping	Specify what should be considered as a peptide (Stripped Sequence, Modified Sequence or Precursor).
Apply Normalization	Uncheck this if you don't want SpectroMine to normalize across the experiment. Choose between median or average global normalization
ILQ Filters	Set two ILQ filters: - Peak-to-Threshold filter – PSM are not used for ILQ quantification if the intensity of the precursor ion divided by noise is lower then set value. - Signal-to-Interference filter - PSM are not used for ILQ quantification if the intensity of the precursor ion divided by summed signal intensity in a whole precursor selection window is lower then set value.
ILQ missing value imputation	Switch on to perform imputation of missing values with the value between 0 and noise values obtained from raw data. Applies only to data obtained with Thermo instruments.

MaxLFQ	Select if the protein level label-free quantification should be performed according to MaxLFQ algorithm (MaxLFQ derives label-free quantities based on inter-run peptide ratios.
XIC Construction	Do XIC construction: check this option if you want SpectroMine to reconstruct XICs to calculate metrics such as apex RT, FWHM and MS1 area under the curve. This is relevant if you want to do DDA label free quantification

PTM Workflow

PTM Localization Filter	Calculates a PTM localization probability for all variable modification site options.
PTM Analysis	Performs a PTM focused data analysis: 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 {REF}. 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. Run clustering specifies whether or not to run hierarchical clustering also on PTM sites.

Minimum localization threshold	Set the filter to chosen level of variable PTMs localization confidence A specified probability cut-off can be applied (default is 0.75).

Post-Analysis

Differential Abundance Grouping	Get the results from differential abundance testing for Protein Group level Smallest quantity unit: the unit to perform t-tests on
Differential Abundance Testing	Choose whether you want SpectroMine to perform testing (paired or unpaired Student's t-test) or not
Assume Equal Variance	Specifies whether the unpaired two sample t-test is run under the assumption of equal variances between the test groups.
Group-Wise Testing Correction	Select if you want to do multi-testing correction by pairing confditions
Gene Ontology	Select controlled vocabulary of gene and gene product attributes as available in the Databases -> GO Databases -> GO ontologies
ILQ efficiency	Check the efficiency of the ILQ labeling and the correct annotation of the channels. In order to check chemical labelling efficiency, ILQ modifications need to be searched as variable and a correct quantification method needs to be selected.
Run Clustering	Choose whether you want to cluster your samples and potential candidates, and how you would like to do that.

Reporting

Select whether you want all the matching proteins to be reported, independently of the result of the protein inference algorithm (IDpicker)

5.2.2 Library Generation

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
Amino Acids	Filter peptides containing specified amino acids
Modifications	Filter peptides according to modifications. Find more details by hovering over the option in the software

iRT Calibration

Set your preferences for iRT calibration:

Auto-assign iRT source	Let SpectroMine keep specific iRT spaces for each relevant source
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).

Minimum Rsquare	Choose how strict you want to be to accept the fit of the iRT calibration of your data

Workflow

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.

If you have a label workflow, itwill create all missing channels, so every peptide will have a light and a heavy counterpart.

5.3 Appendix 3. Analysis Perspective Plots

5.3.1 Run Node Plots

5.3.1.1 RT Calibration Plot

This plot shows the iRT calibration for the selected run. SpectroMine™ supports precision, non-linear iRT calibration using a large set of iRT peptides contained in an internal database (Box 3).

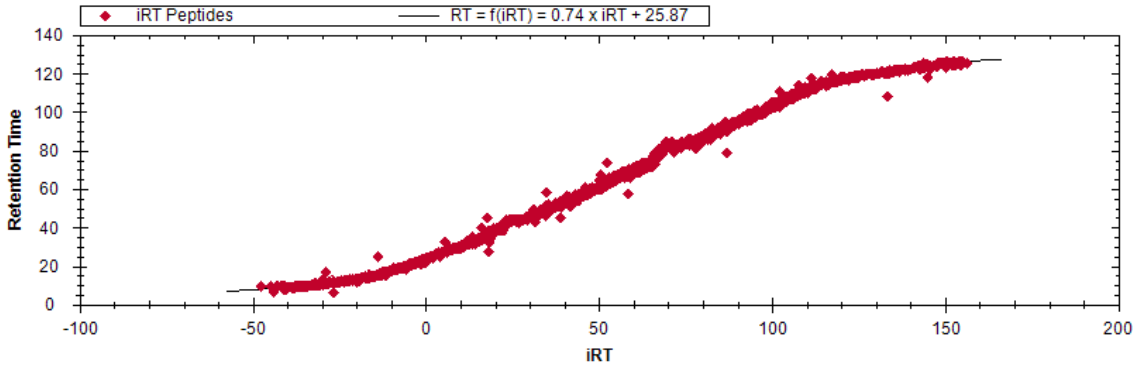

Figure 43. iRT calibration plot showing the non-linear transformation from iRT peptide repository to actual predicted retention times. The extended calibration allows you to correct even small gradient fine-structure fluctuations in order to get the most accurate retention time prediction for your run.

5.3.1.2 TIC Chromatogram & TIC Overlay Chromatogram

The TIC Chromatogram plots show your run's total ion current chromatogram. The TIC Overlay Chromatogram combines the information of all runs of your experiment in one plot for better insight into instrument stability and amount of sample injected.

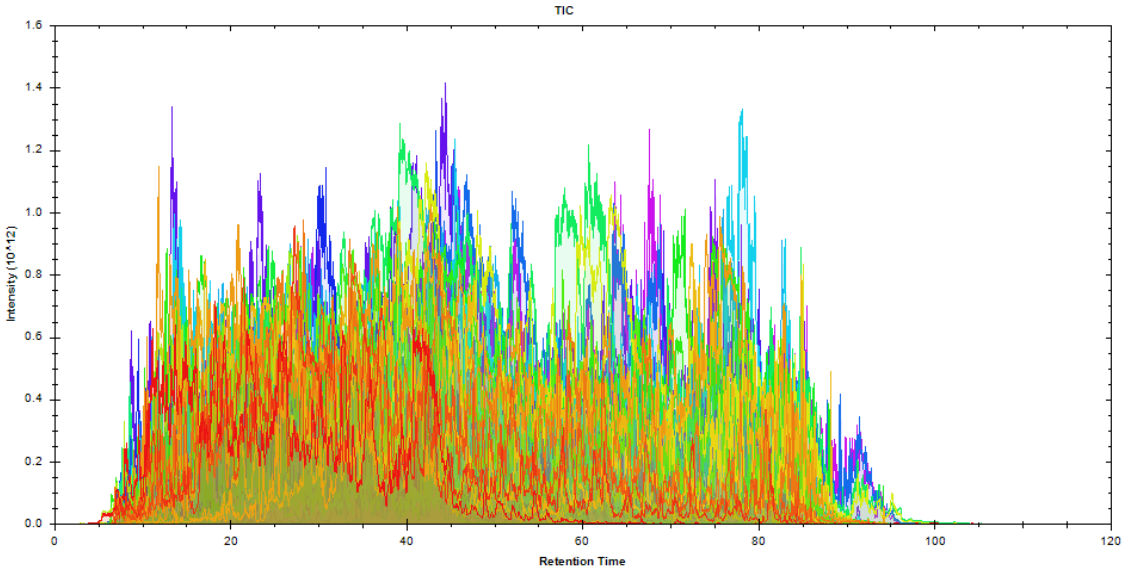

Figure 44. The TIC Overlay chromatogram shows the total ion current of all runs in your experiment in a single plot, giving insight into instrument stability and amount of sample injected. In the example shown, TICs come from fractionated samples.

5.3.1.3 MS1, MS2 and MS3 Mass Calibration

The mass calibration plots show the result from the calibration performed by SpectroMine at MS1, MS2 and MS3 (if applicable) levels as a function of retention time of precursors and as a function of their m/z values. You can see the mass error for each feature before and after calibration.

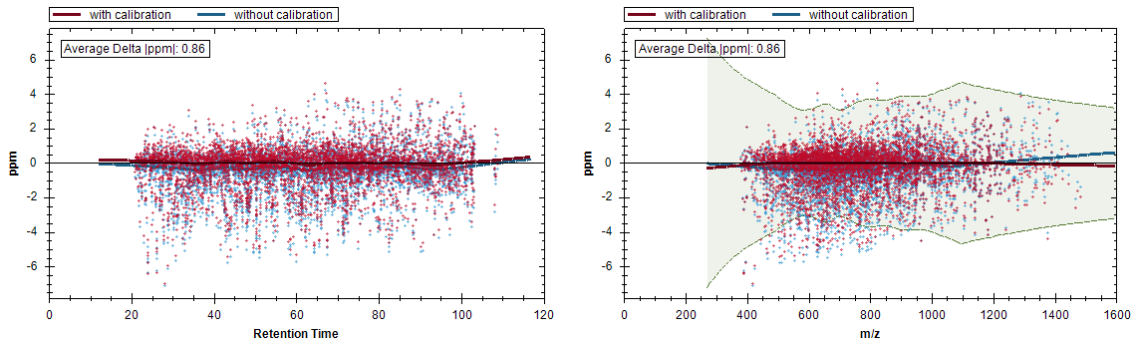

Figure 45. Mass calibration plots in the function of precursor retention time (left side panel) and in the function of precursor m/z value (right side panel). SpectroMine performs mass calibration to optimize the search accuracy. The blue spots correspond to the mass error of the corresponding level features (M1, MS2 or MS3) before SpectroMine calibration. The red spots show the result of the calibration.

5.3.1.4 Score Histograms

Under this plot selection, SpectroMine shows the behavior of the target and decoy distributions for each of the round searches performed, in different tabs.

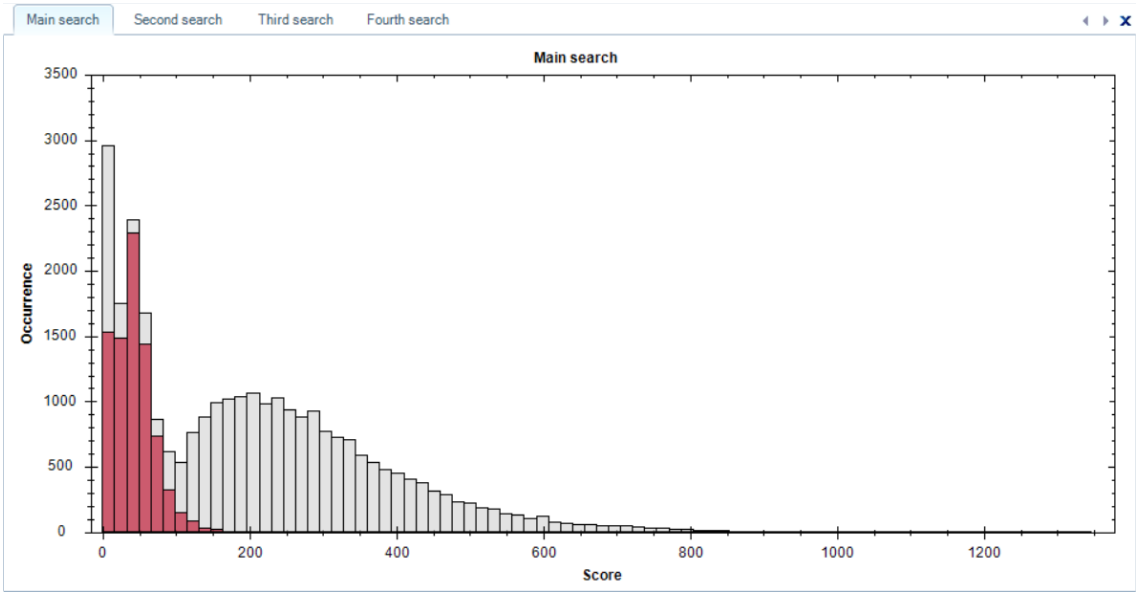

Figure 46. Behavior of the target (gray) and decoy (red) distributions for each of the round searches performed (in tabs).

5.3.1.5 PSMs per Search

The PSMs per search plot depicts the number of peptide spectrum matches (PSMs) identified in each search performed.

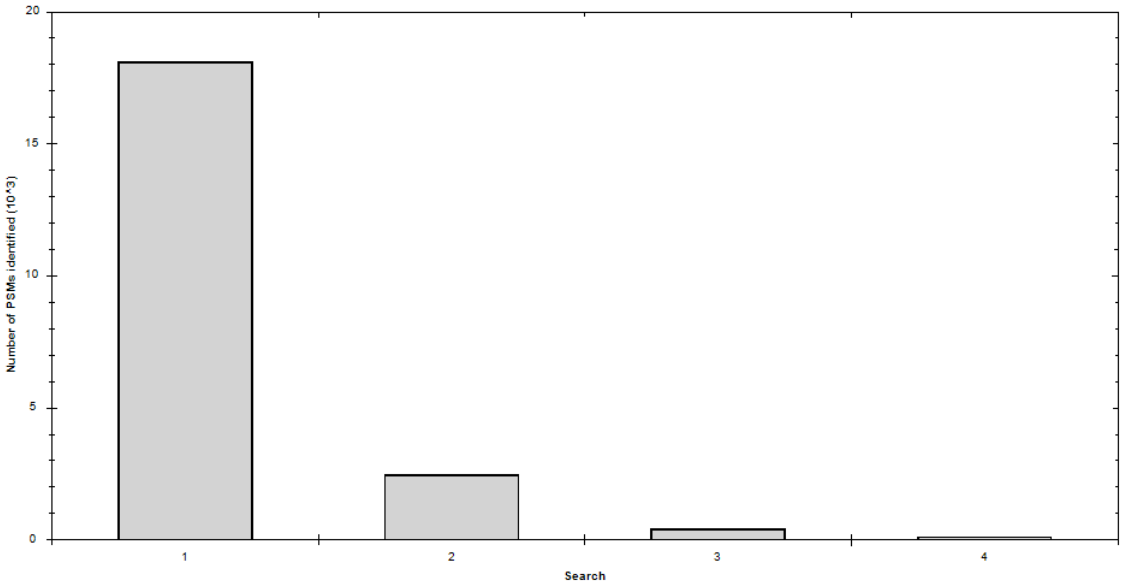

Figure 47. Number of PSMs identified in each round search.

5.3.1.6 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.

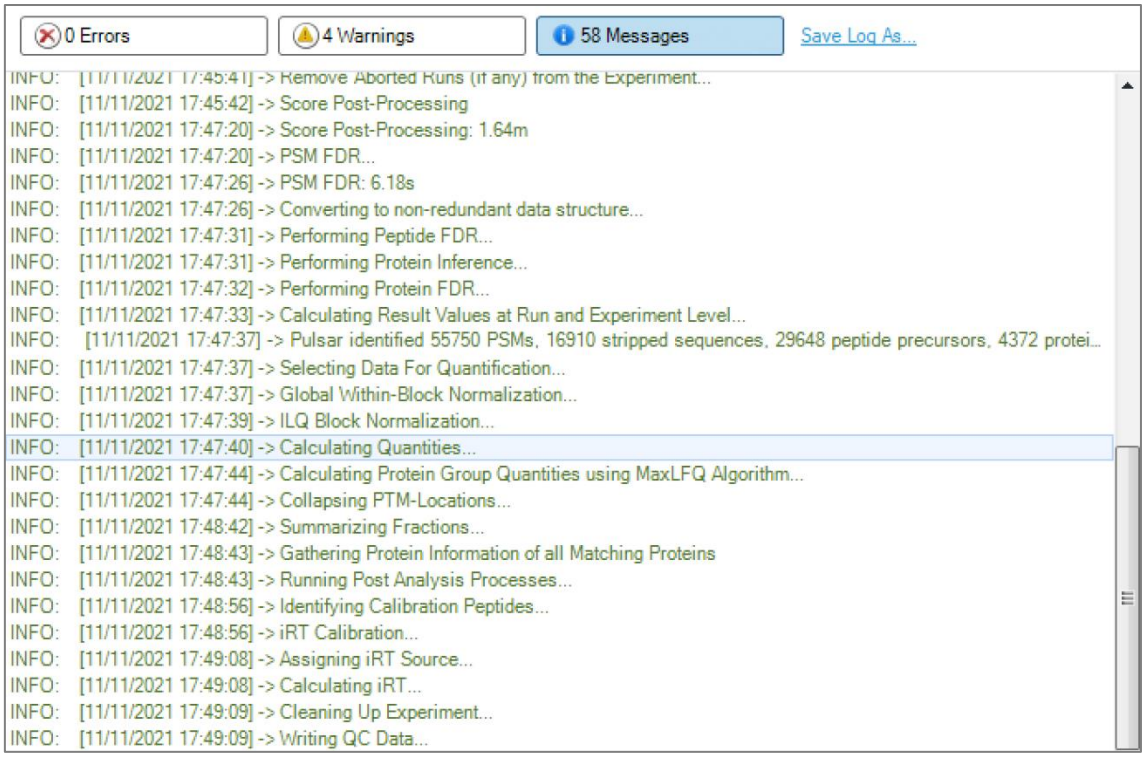

Figure 48. The analysis log with detailed information about the analysis processes in SpectroMine. The particular tabs will report errors, warnings and informational messages related to the analysis.

5.3.2 Protein Level Plots

5.3.2.1 Protein Coverage and Protein Inference Information

SpectroMine gives you a detailed overview of a protein's coverage and the FASTA file information corresponding to that inference. The protein coverage plot shows you all the peptides of a protein that were identified. You can choose to show the coverage details on experiment level by selecting it upon right-clicking the plot.

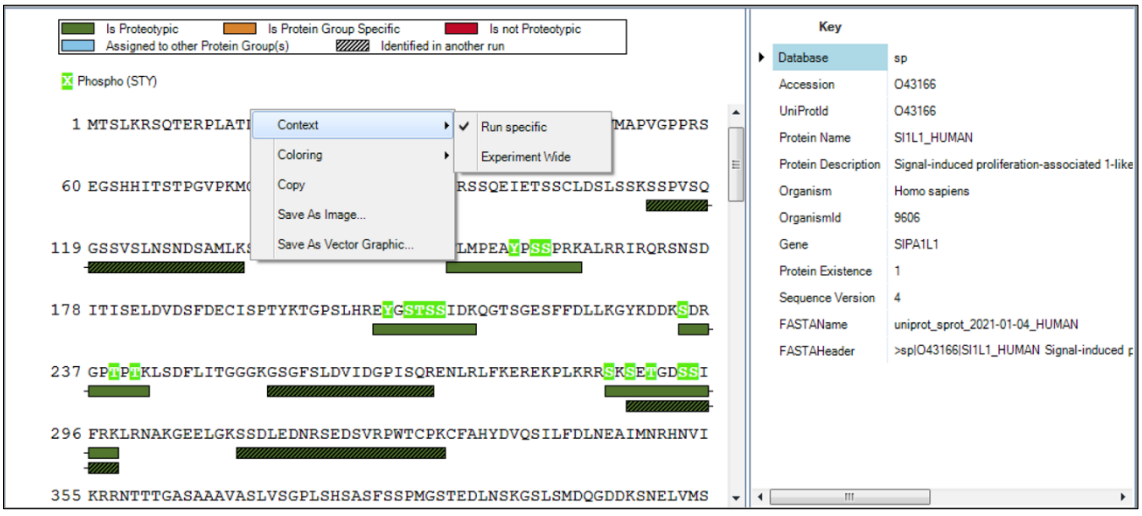

Figure 49. Protein coverage plot and database information for protein O43166. You can choose to show the coverage details on experiment level by selecting it upon right-clicking the plot. Confidently localized PTMs are highlighted above identified peptides.

5.3.2.2 Protein Group Quantity Plot

You can visualize the intensity of a given protein group across runs (for the LFQ experiments) or across the channels (for the ILQ experiments). The exact quantity in each run of the LFQ experiment (or channel for the ILQ experiments) can be annotated by right-clicking on the bar plot.

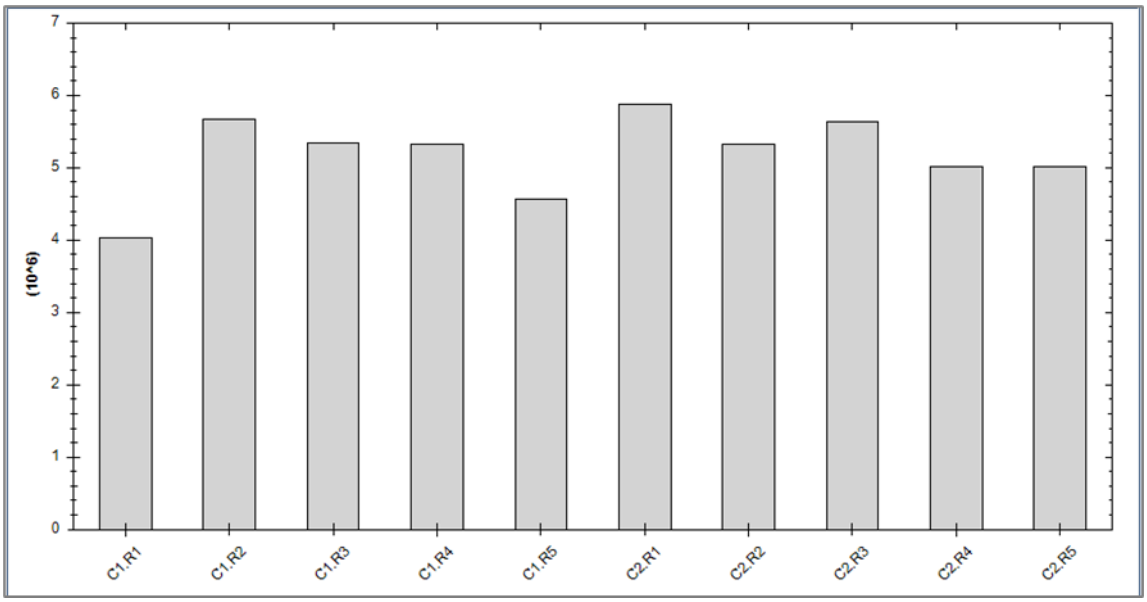

Figure 50. The Protein Group Quantity plot visualizes the levels of a given protein across labeled channels in ILQ experiment.

5.3.3 Precursor Level Plots

At the precursor level you can inspect precursor summary, which among other precursor related features, indicates number of PSMs matched for that precursor. You can visualize Best PSM plot, Corresponding MS3 spectrum and MS1 Isotop Envelope Extraction (of all PSMs). All of those graphs are described in detail on the next section related to the PSM level Plots.

Moreover, at the precursor level you can inspect Precursor Quantity described below.

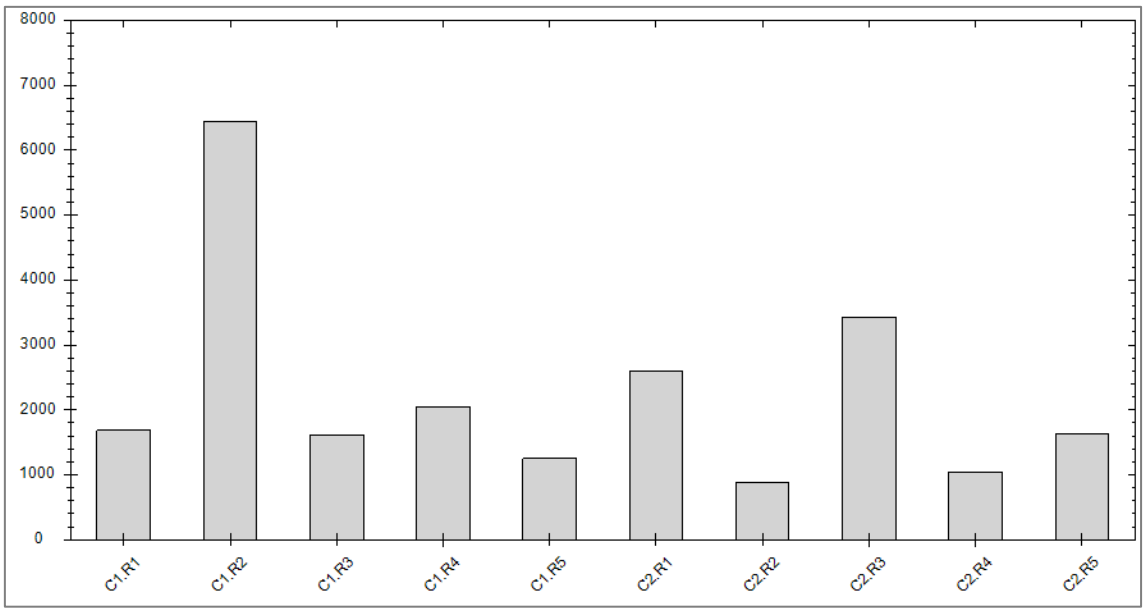

Figure 51 The Precursor Quantity plot visualizes the levels of a chosen precursor in each labeled channel.

5.3.4 PSM Level Plots

5.3.4.1 PSM Plot

The PSM plot shows peptide-spectrum matches (PSMs) for an identified precursor. 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 detected fragments for the peptide you are looking at. 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 is depicted. 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 diagnostic peaks (d) or reporter ions (r) on spectrum
- Show the mass error in ppm or in Th.
- Choose which fragments to highlight on the MS2 spectrum: the ones identified and used for scoring, all the matched ones, or all the theoretical fragments. You can also filter by fragment length, by type and by neutral loss.

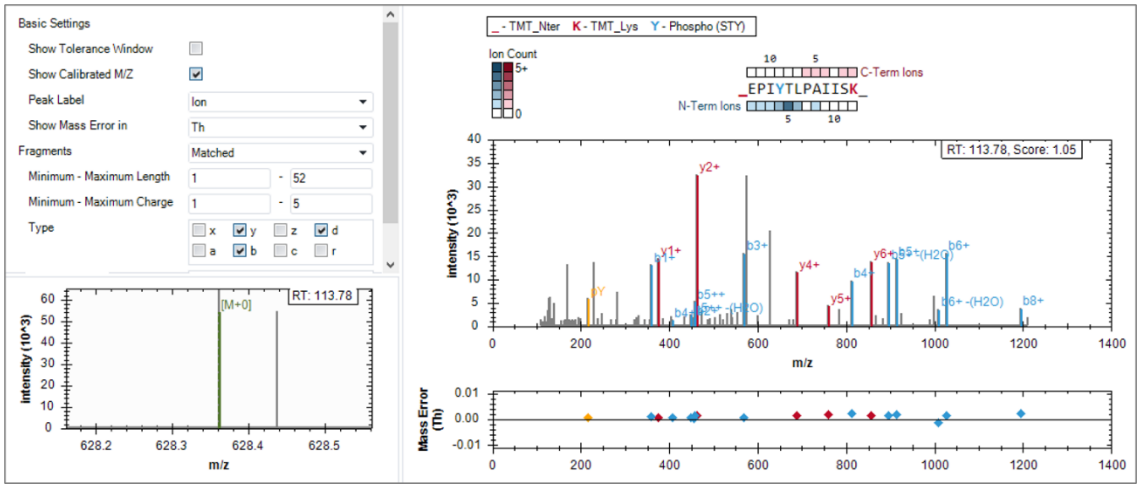

Figure 52. PSM plot for a particular precursor. At the top-left panel you can select what to highlight on the MS2 spectrum plots. In this case, all the matched fragments fulfilling the filter criteria where shown. In blue you can see the b-ions: in red, the y-ions. If you selected additionally diagnostic peaks ions (d), they will be shown in yellow. The mass errors, in ppm, are shown for each fragment. The MS1 spectrum is shown at the bottom-left.

5.3.4.2 Corresponding MS3 spectrum

If there is MS3 information, you can see the MS3 spectrum in this plot. This plot is similar to the PSM plot (Figure 52) but on MS3 level. In a SPS-MS3 quantification experiment, the reporter ions can be nicely observed, like in the example below (Figure 53).

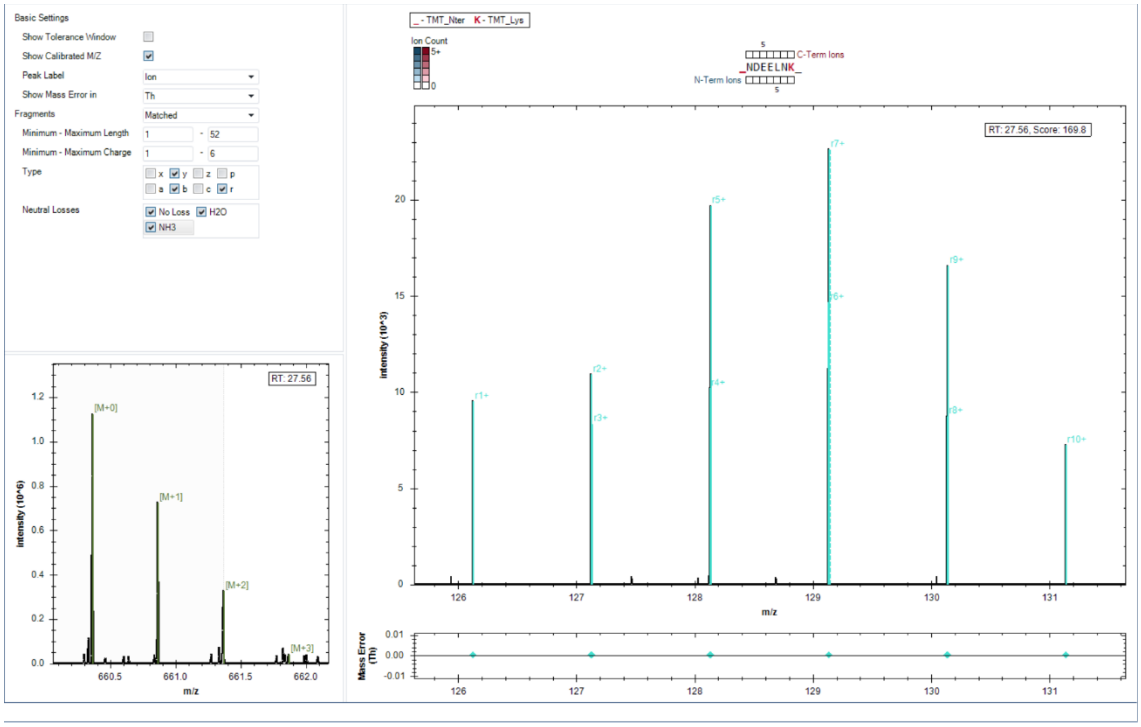

Figure 53. Corresponding MS3 spectrum. This plot shows the MS3 spectrum match, if applicable. In this example, the reporter ions for a TMT10-plex experiment can be nicely observed for a particular precursor.

5.3.4.3 MS1 Isotope Envelope XIC

This plot shows the monoisotopic precursor plus its most abundant isotopic forms as an XIC chromatogram. The MS1 XIC chromatogram is color coded to reflect the predicted relative intensities. A color coding from red (highest) to blue (lowest) indicates high correlation with the predicted abundance.

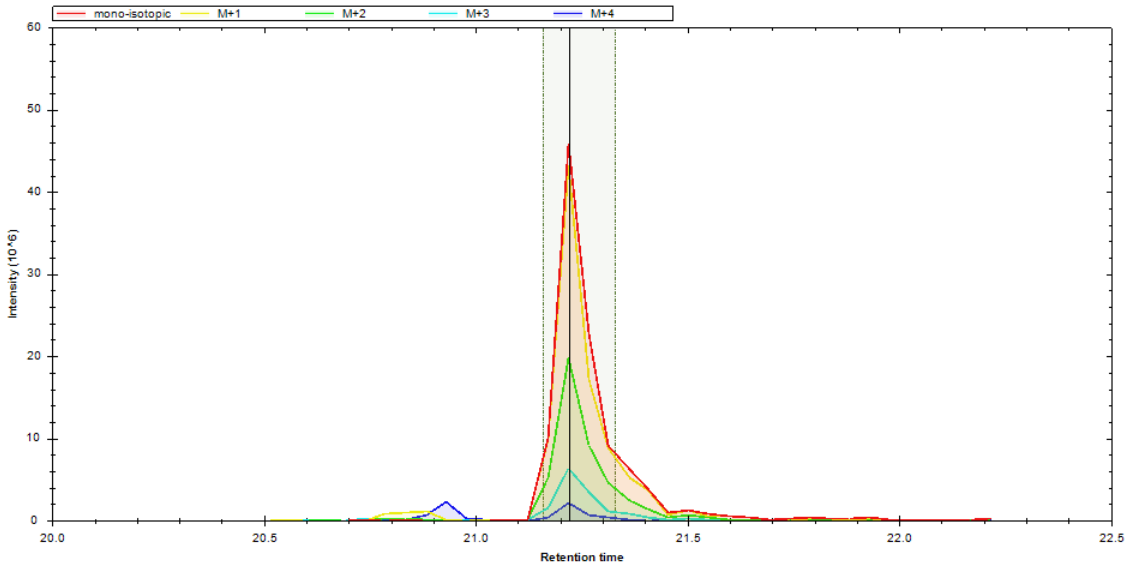

Figure 54. The MS1 Isotope Envelope XIC chromatogram for the precursor as extracted by SpectroMine. The color pattern again indicates a high correlation to the expected relative abundances.

5.3.4.4 Fragment Intensity Prediction

This plot shows predicted fragment intensity of a particular precursor as well as empirical data obtained in the experiment. After right-clicking you can change the scale to the logarithmized values.

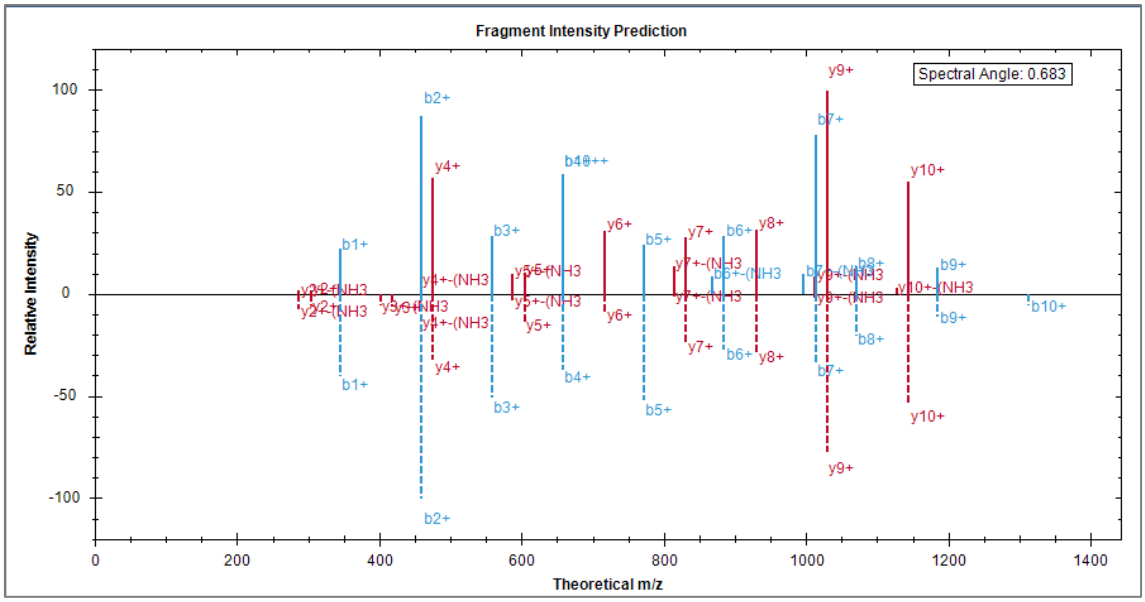

Figure 55. Fragment Intensity Prediction plot visualizes relative instensity of signals measured for a given precursor (upper panel) and the predicted spectra for that precursor (lower panel) in a function of theoretical m/z.

5.4 Appendix 4. Post Analysis Perspective Plots

The Post Analysis Perspective has plots that help visualize results of the experiment. Among other, there is a number of graphs that help to assess the quality of obtained data – Analysis Overview as well as Scoring Histograms. Examples of graphs belonging to those two categories are presented below. More graphs can be found in the main Post Analysis Perspective section 3.5.

5.4.1 Modification Type

Modifications Per Type bar plot shows the number of peptides with assigned fixed and variable modifications across the whole experiment.

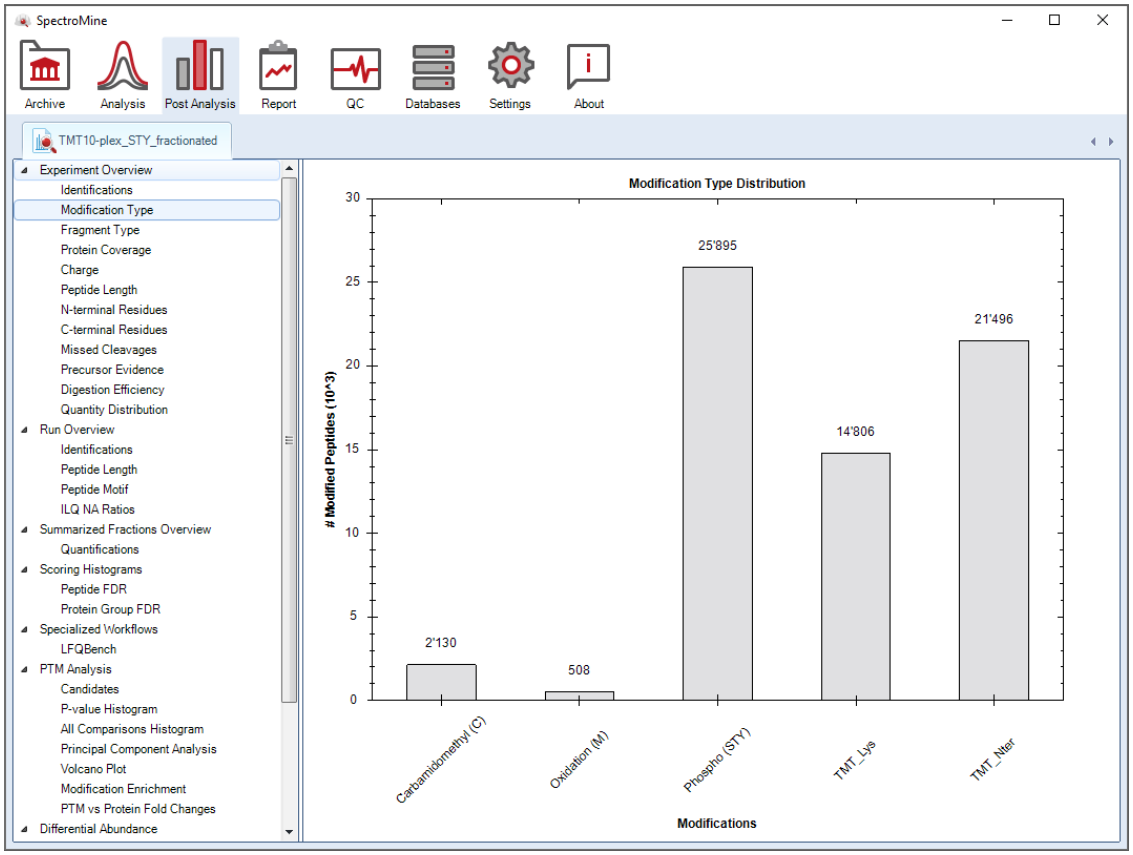

Figure 56 Modifications Per Type bar plot shows the total number of modified peptides per each modification across whole experimental set.

5.4.2 Fragment Type

5.4.3 Protein Coverage

The Protein Coverage plot shows the distribution of identified precursors per protein across the whole experiment.

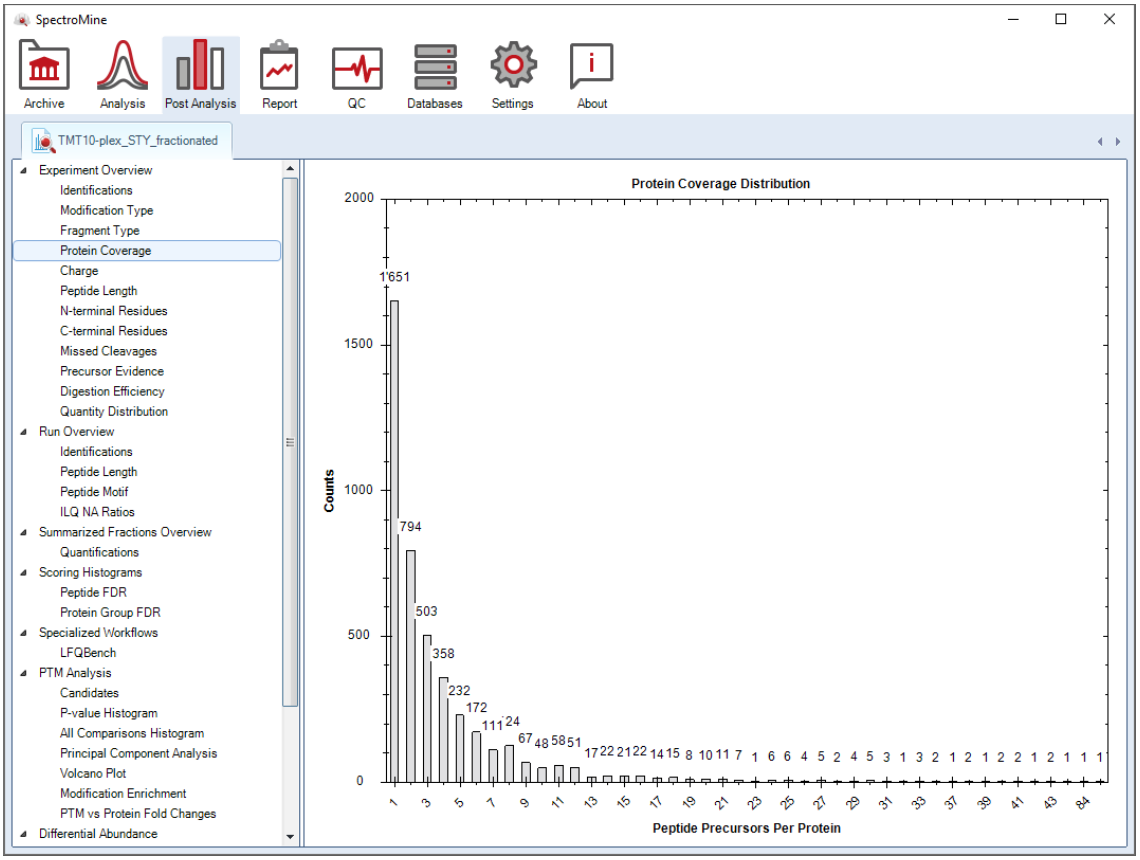

Figure 57. The protein Coverage shows distribution of numbers of precursors per protein.

5.4.4 Quantity Distribution

The Quantity Distribution plot shows box plots with the distribution of quantity for each of the reporter ions across the whole experiment. For the visualization, also logarithmized quantity values could be chosen.

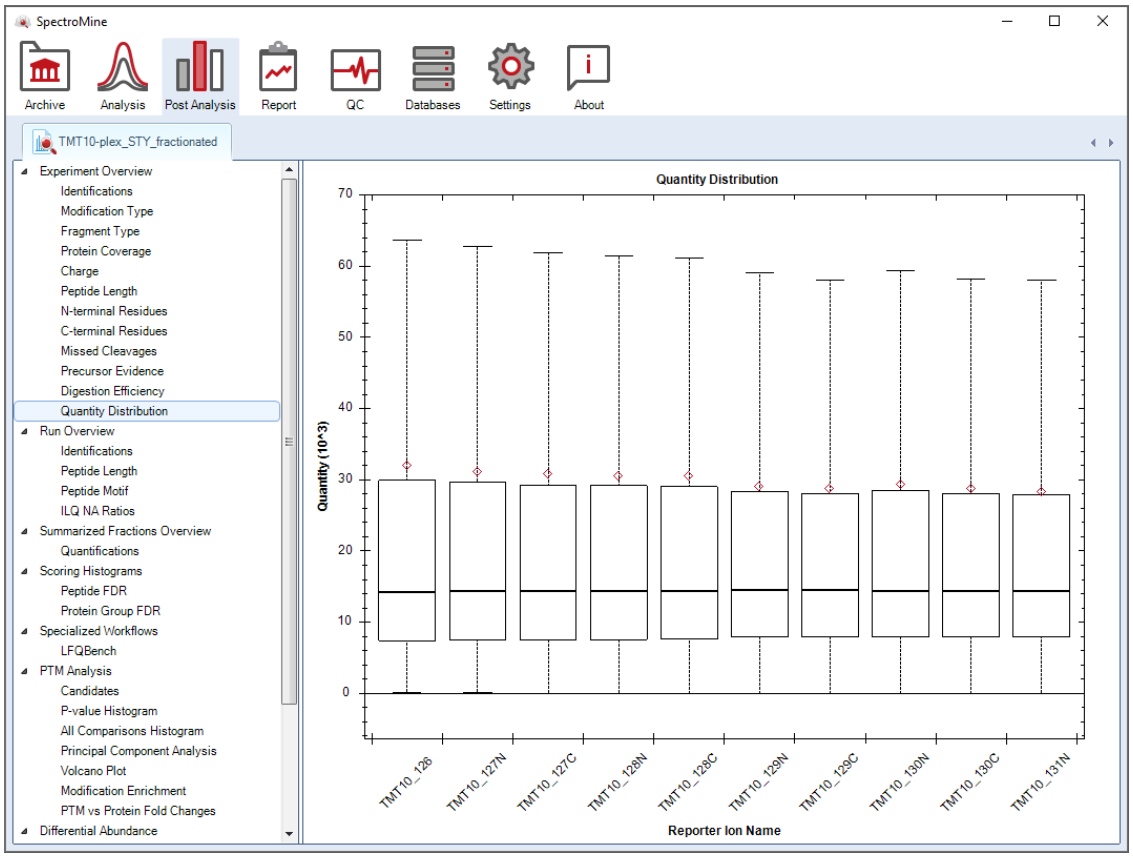

Figure 58. The Quantity Distribution Plot shows the distribution of the reporter ions used for the quantification in the experiment.

5.4.5 Identifications (Run Overview)

The Identifications plot in the Run Overview presents the number of identifications per run displayed in a bar plot (Figure 59). It also helps to investigate the cumulative sparse and full identifications in the experiment. The total number of cumulative sparse identifications should increase or stay at the same level after adding information from each run to the next. For the cumulative full identifications the opposite trend is usually observed since the overlapping identifications tend to decrease by adding more runs and complete profiles are harder to obtain. On right-click you can change the basis of quantification between precursor, modified sequence, stripped sequence and protein group.

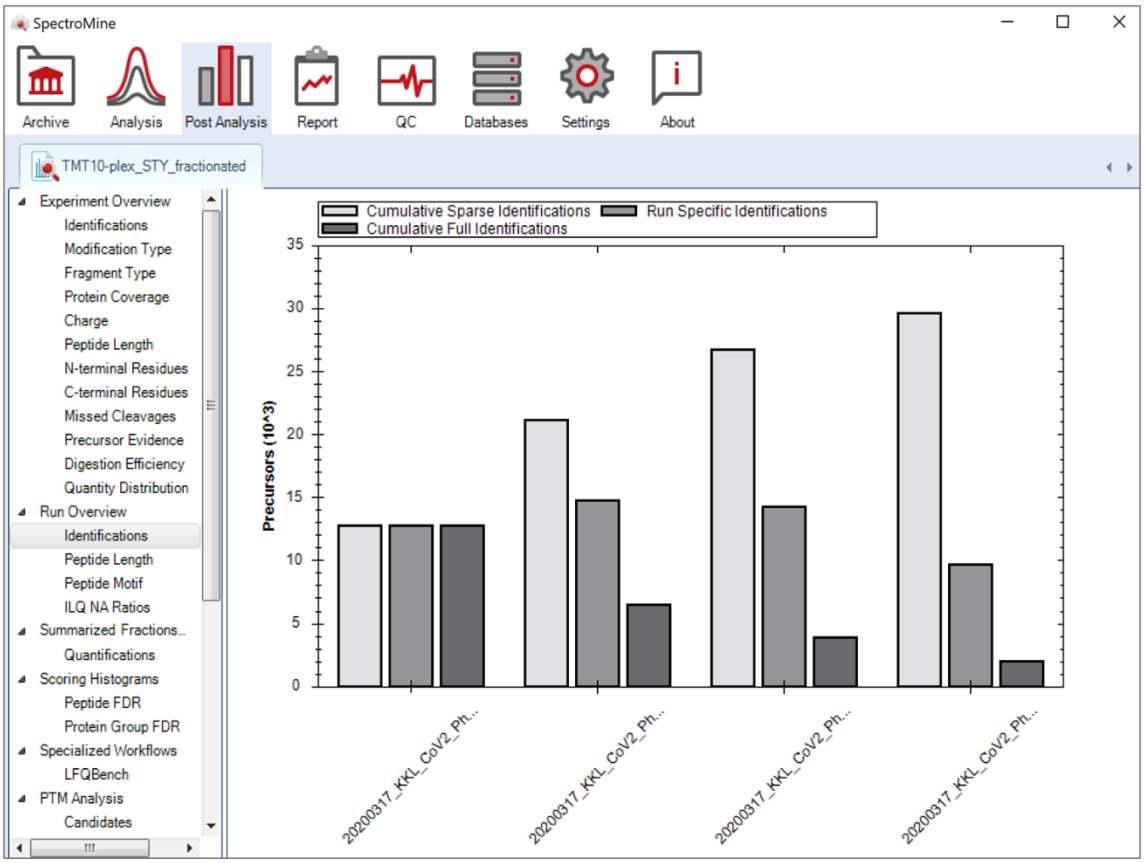

Figure 59 Identifications per run bar plot shows the number of identifications per run (middle gray), the cumulative full (dark gray) and sparse identifications (light gray).

5.4.6 Peptide Motif

The Peptide Motif plot is conceived for immunopeptidomics workflows and highlights the enriched peptide sequences in a sample.

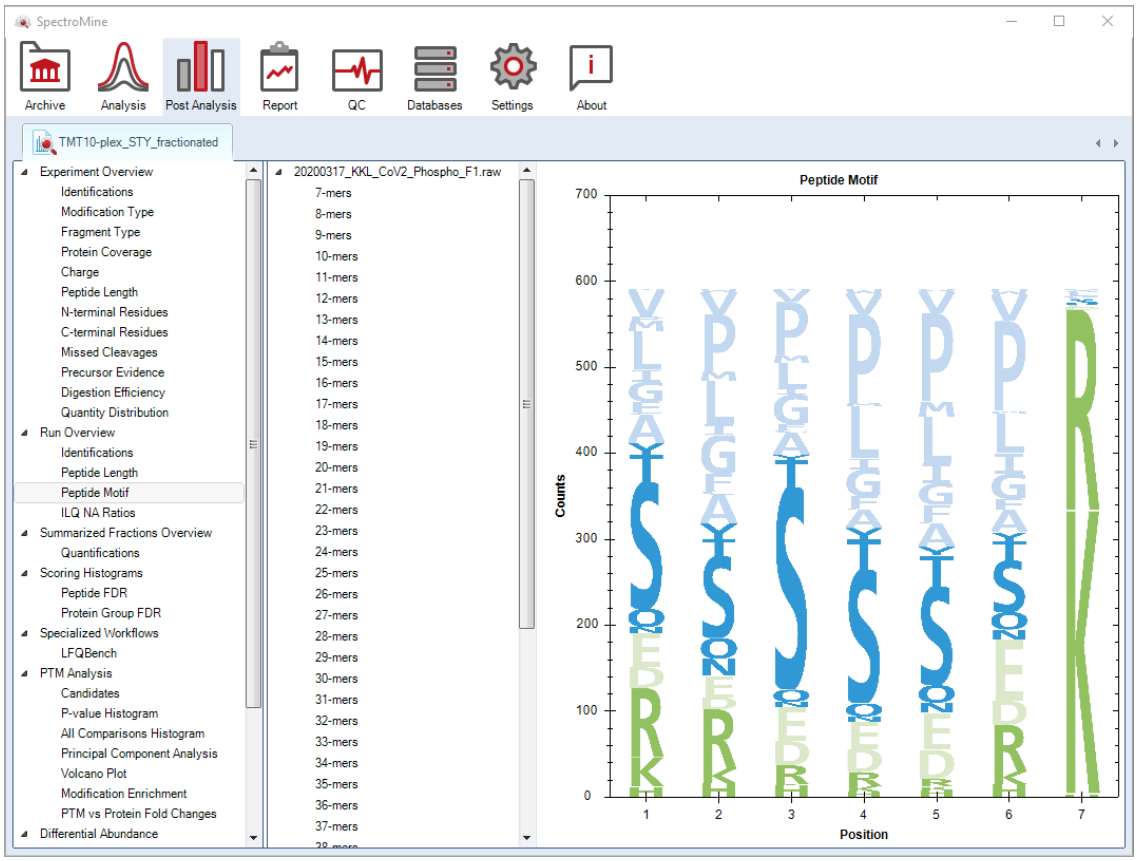

Figure 60. The Peptide Motif plot shows for each n-mer within a run, the amino acid counts per position.

5.4.7 Peptide FDR

The Peptide FDR plot shows the scores distribution for target and decoy peptides. The number of bins in the histogram can be set by the user.

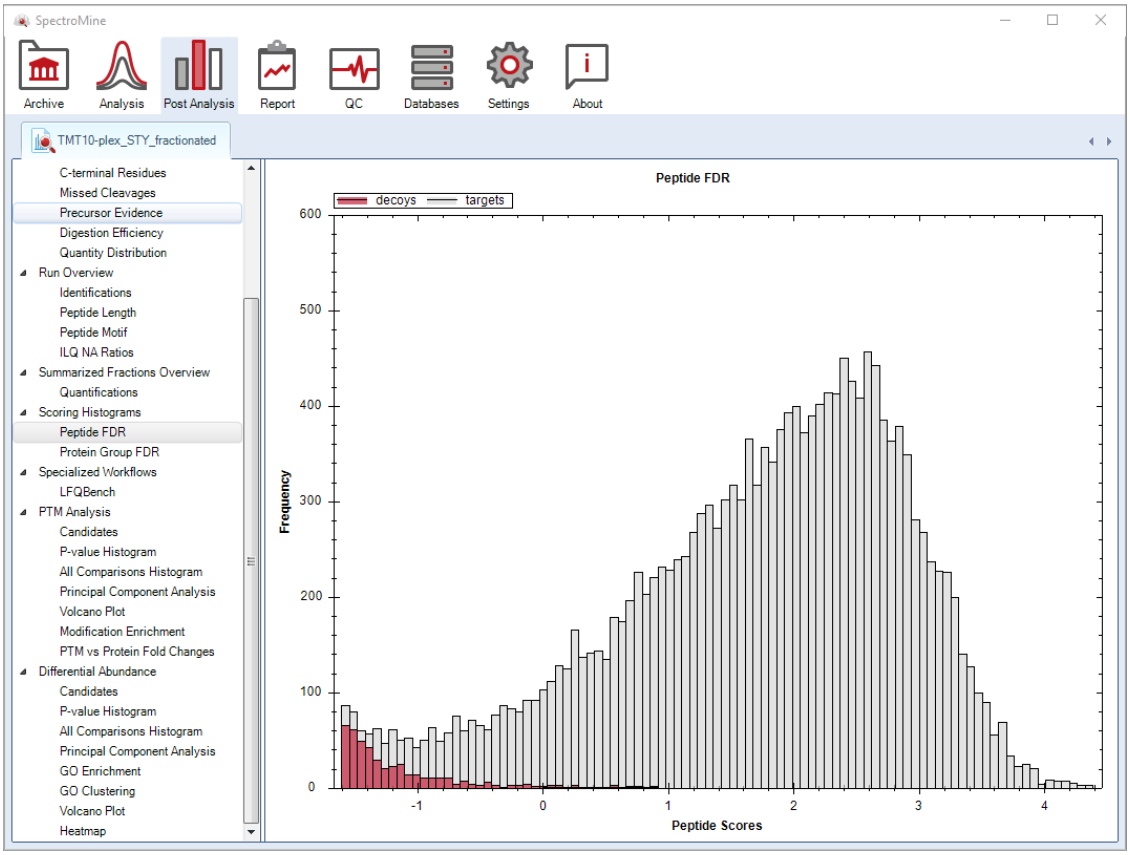

Figure 61. The Peptide FDR plot shows the distribution of the peptide target (light gray) and decoy (red) population.

5.4.8 Protein Group FDR

The Protein Group FDR plot shows the distribution of protein group scores for the identified targets and decoys. Number of the bins in histogram could be set by the user.

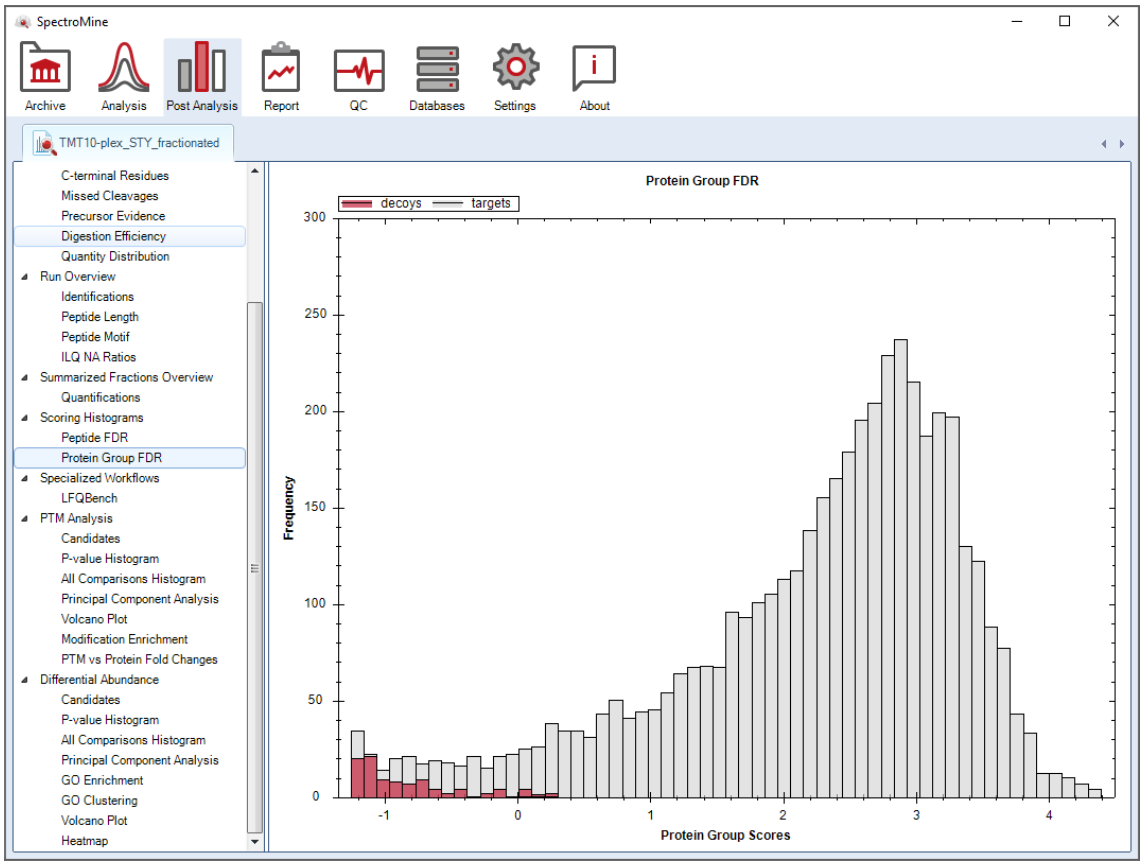

Figure 62. The Protein Group FDR plot shows the distribution of scores at the protein group level for target protein groups (in light gray) and decoys (in red).