

Tutorial 9: SWATH data analysis in Skyline

In this tutorial we will learn how to perform targeted post-acquisition analysis for protein identification and quantitation using a data-independent dataset acquired in a QqTOF instrument (5600 TripleTOF, AB Sciex).

The data come from a *Mycobacterium tuberculosis* experiment where a mutant has been evolved which is drug resistant. Our hypothesis is that some proteins responsible for maintaining iron homeostasis may be affected by this mutation. As such we will focus on proteins of the IdeR regulon. The number of peptides and proteins which we will extract is relatively small, however, keep in mind that different sets of functional groups of proteins could be extracted (or, potentially, the entire proteome) from the same data set.

Initially, we will set all the parameters in the Skyline session required to work with data-independent datasets and then we will proceed to extract the quantitation information from the raw data files.

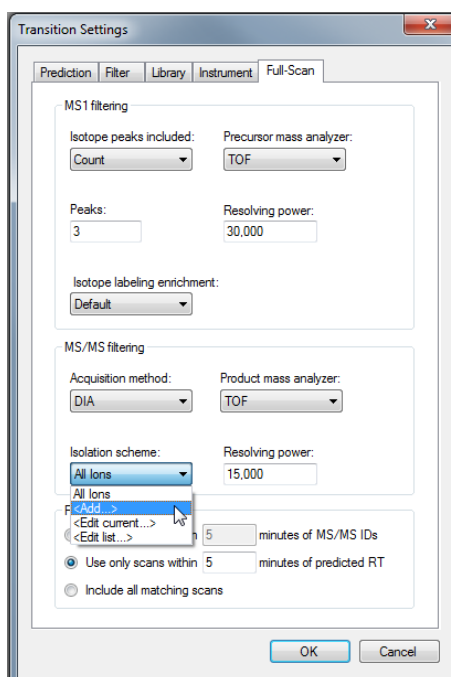
1. Preparing Skyline session for data-independent acquisition (SWATH)

In this section we are going to prepare a Skyline session with the appropriate settings for SWATH analysis. We will tune the settings for data-independent analysis, and then we will extract our proteins of interest from the SWATH file.

- Open the Skyline document *TarProCourse_Tut9_swath.sky* from the folder *Tutorial-9_SWATH*
Note! In the “Targets window” there are the iRT peptides already listed and later we will populate the window with additional targets.

1.1. Defining the data-independent acquisition isolation scheme

- Go to “Settings” → “Transition Settings” → “Instrument”
 - Change the “Max m/z” value to 2000 m/z
- Go to “Full-Scan”
Fill the window as indicated in the screen shot.
Note! The resolving power depends on the state of the instrument during SWATH data acquisition. The optimum might be slightly different for each dataset.



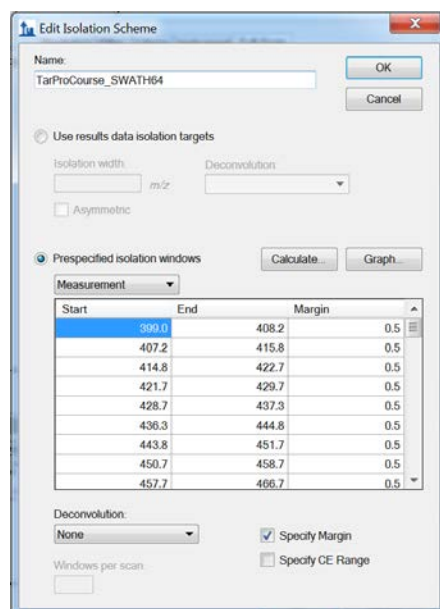
Now we need to define a new “Isolation scheme” according to the parameters defined on the instrument for data-independent acquisition.

Note! In our case we used 64 variable width windows that covered the range from 400 to 1200 m/z with 1 m/z overlap.

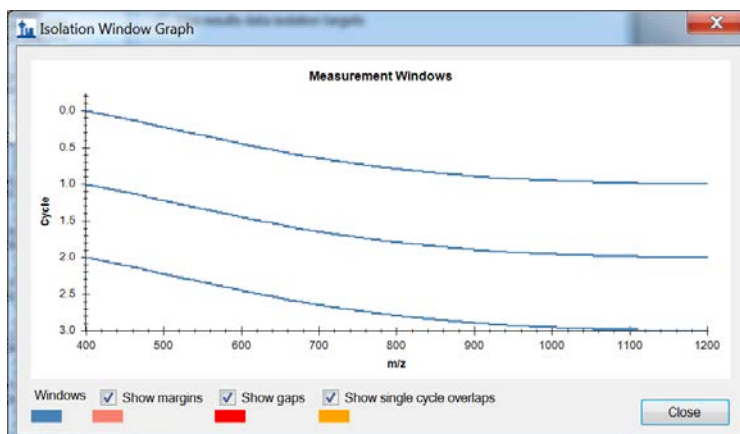
- From “Isolation scheme” drop-down → select “Add...” → fill the name as “TarProCourse_SWATH64”
- Select “Prespecified isolation windows” and activate “Specify Margins”
- Go to your “Tutorial-9_SWATH” folder and open the file “bamir_K151128_009_SW_swathWindows.csv” in Excel. The first column of the table is the start and the second column is the end m/z of each isolation window.
- Add a third column specifying the margin = 0.5 for all rows.

Note! As the transmission windows are not perfect, the margin option allows to specify how much of the window edges should not be used for extraction. Skyline will then extract from *start+margin* until *end-margin*

- Copy the excel table with *ctrl-c* and paste it with *ctrl-v* into the Skyline table for isolation windows.
- The “Edit Isolation Scheme” window should now look like this:



- Click “Graph...” to see how the isolation windows cover the specified range.



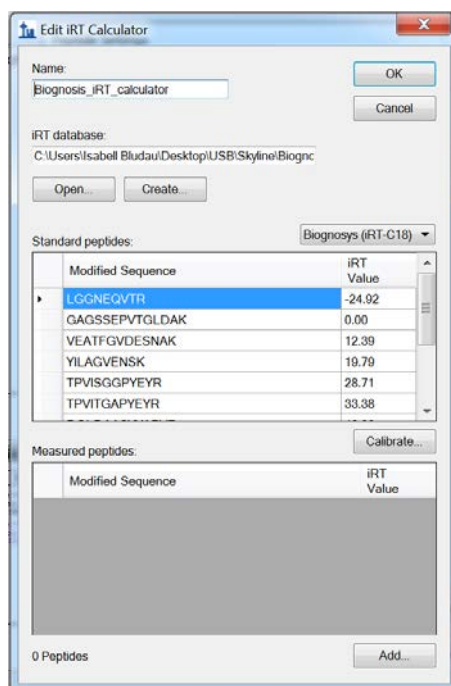
- Click “Close”
- Click “OK” twice
- Back in in the “Full-Scan” tab, make sure that the “Retention time filtering” is set to “Use only scans within 5 minutes predicted RT”. This is used for setting a 10 minutes window around the predicted RT.
- Click “OK”.
- Save your skyline document as *TarProCourse_Tut9_swath_acquisitionscheme.sky*

1.2. Adding an iRT calculator

- Go to “Settings” → “Peptide Settings” → “Prediction”
- Click on the calculator symbol
 - Select “Add...”
 - Name: “Biognosis_iRT_calculator”
 - iRT database: click “Create...” → Select your Skyline folder and save the file with the name: “Biognosis_iRT_calculator”
 - From the dropdown menu on the right select “Biognosis (iRT-C18)”

NOTE! Skyline offers the direct application of some standard iRT kits including the Biognosis kit that was used for this experiment. You will see the iRT peptides from the Biognosis iRT kit appearing in the standard peptides table.

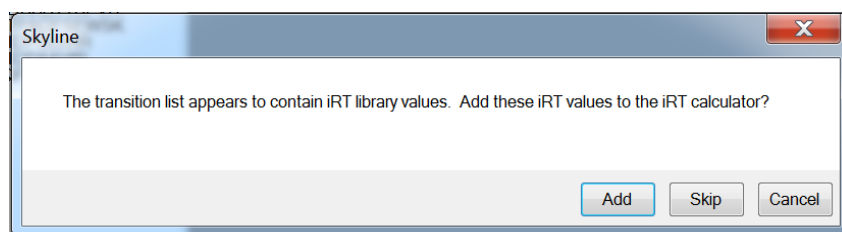
- The window should now look like this:



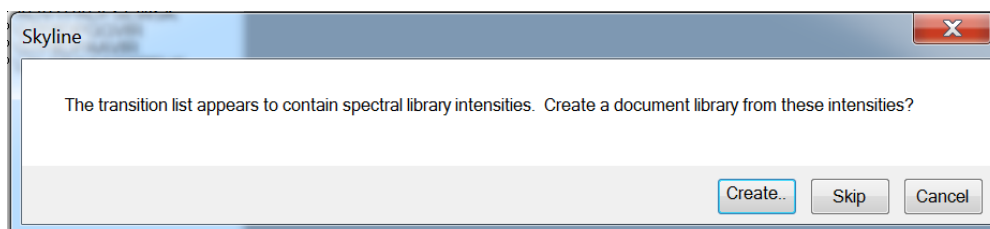
- Click “OK”
- Back in the prediction tab of the peptide settings go to the dropdown menu for the retention time predictor and select “Add...”
 - Name: “Biognosis_iRT_predictor”
 - Select “Auto-calculate regression”
 - Set the “Time window” to 4 min
 - From the calculator dropdown menu choose your new “Biognosis_iRT_calculator”
 - Click “OK”
- Click “OK”

1.3. Adding an assay library

- The assay library we will use is a subset from a published proteome-wide SWATH assay library for Mtb (1. Schubert, O. T., Ludwig, C., et al. Absolute Proteome Composition and Dynamics during Dormancy and Resuscitation of Mycobacterium tuberculosis. Cell Host & Microbe 18, 96–108 (2015)). The complete library is available from www.swathatlas.org along with several other published SWATH assay libraries from different organisms. The library has been created according to a published protocol (Schubert, O. T. et al. Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat. Protocols 10, 426–441 (2015)). We have made a subset of the published assay library containing the 3 top flying peptides for 40 proteins belonging to the IdeR regulon responsible for maintaining iron homeostasis in Mtb.
- Open the file “*Mtb_IdeR_bestPeptides_library_64w.csv*” in Excel. Notice the headers and format of this file and the type of information contained in the assay library. Close Excel.
- Go to “File” → “Import” → “Transition List...” and select “*Mtb_IdeR_bestPeptides_library_64w.csv*”
- Window popping up: click “Add”



- Window popping up: click “Create..”



- Now your assay library is being uploaded, you can follow the process on the progress bar.
- Reopen the iRT calculator (Settings → Peptide Settings; click the calculator symbol and select ‘Edit Current’). Notice that the ‘Measured peptides’ window had been populated with the target peptides and their corresponding iRT values from the assay library.
- Open the Spectral Library Explorer (View → Spectral Libraries). Note that pseudo MS/MS spectra have been created for using the relative fragment ion intensities read from the assay library. Each pseudo-spectrum contains 6 fragments only because this is what has been selected for this assay library.

NOTE! The assay library only contains target proteins, thus decoys need to be generated in the next step.

- Go to “Edit” → “Refine” → “Add Decoy Peptides...”
- Leave the default number of decoy precursors.
- Select “Reverse Sequence” from the decoy generation method dropdown menu.

- Click “OK”
- Save your skyline document as *TarProCourse_Tut9_swath_assay_library.sky*

2. Performing the SWATH data analysis

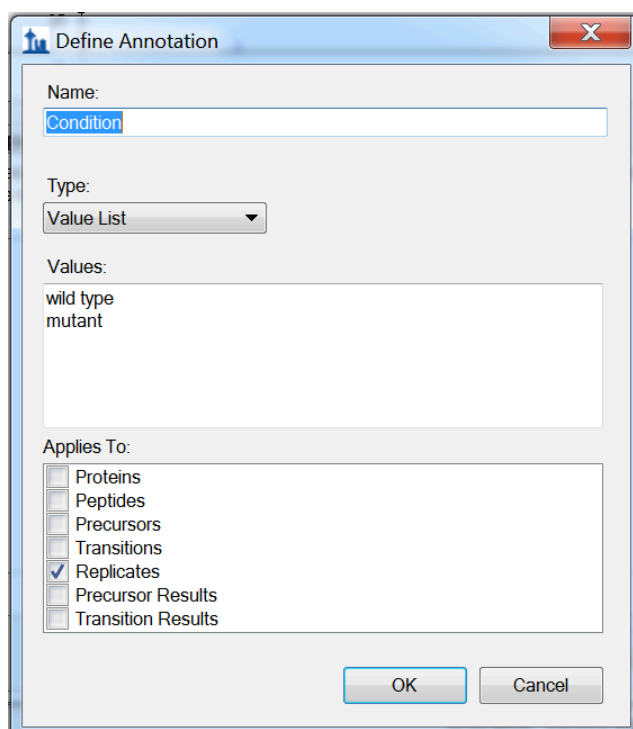
2.1. Import and extraction of SWATH data

- Go to “File” → “Import” → “Results...”
- Select “Add single-injection replicates in files”
- Select “Show chromatograms while loading”
- Select all 6 SWATH files:
 - bamir_K151128_001_SW_qtofpp.mzXML
 - bamir_K151128_004_SW_qtofpp.mzXML
 - bamir_K151128_005_SW_qtofpp.mzXML
 - bamir_K151128_008_SW_qtofpp.mzXML
 - bamir_K151128_009_SW_qtofpp.mzXML
 - bamir_K151128_012_SW_qtofpp.mzXML
- Click “OK”

NOTE! The SWATH data gets imported now and target and decoy transitions are extracted. This process can take some time. For this tutorial we selected data that was already converted from the raw data format to mzXML and centroiding was performed (using the qtofpeakpicker). Using centroided data increases processing speed compared to profile data. Furthermore we use a small assay library specific for proteins involved in IdeR pathway. In a lot of applications larger assay libraries are chosen resulting in a longer time for the data extraction process.

2.2. Sample Annotation

- In the next step we want to annotate the samples.
- Go to “Settings” → “Document Settings” → “Annotations”
- Click “Edit List...” and select “Condition” → click “Edit...”
- Modify the values in the “Define Annotation” window as follows:



- Click “OK” twice
- Back in the “Document Settings” select **Condition** and **BioReplicate**
- Click “OK”
- Go to “View” → “Results Grid”
- Annotate the samples as shown in the figure below:

Results Grid

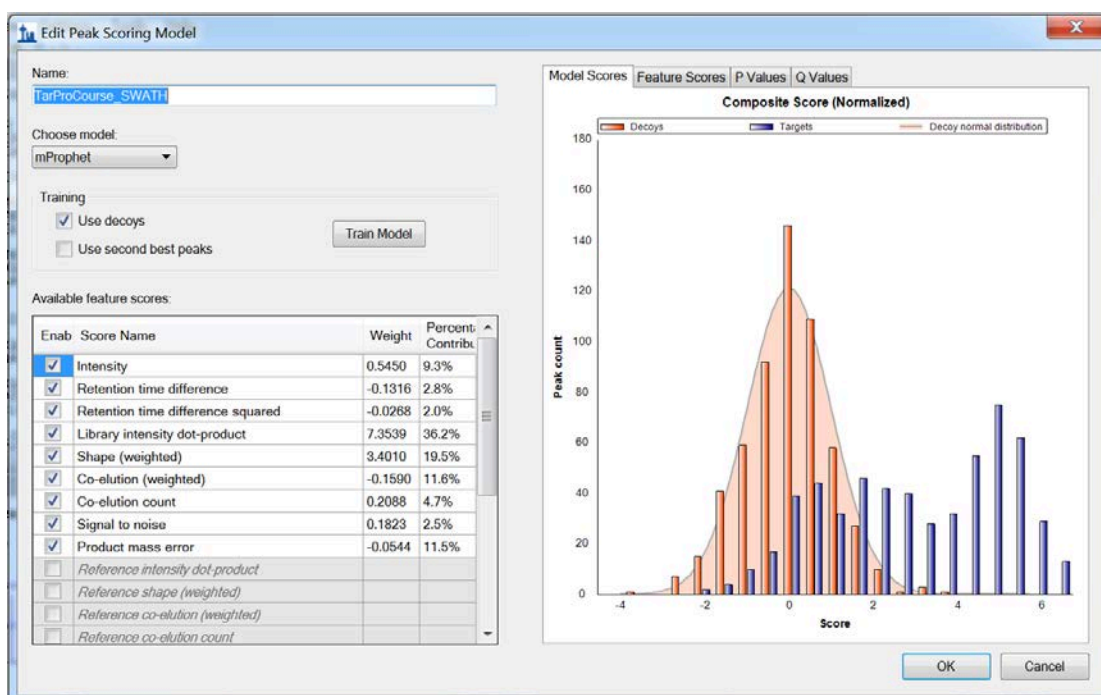
Views 6 of 6 Export... Find:

Replicate	Sample Type	Analyte Concentration	Condition	BioReplicate
001 SW ...	Unkno...		wild ty...	1
004 SW ...	Unkno...		mutant	1
005 SW ...	Unkno...		wild ty...	2
008 SW ...	Unkno...		mutant	2
009 SW ...	Unkno...		wild ty...	3
012 SW ...	Unkno...		mutant	3

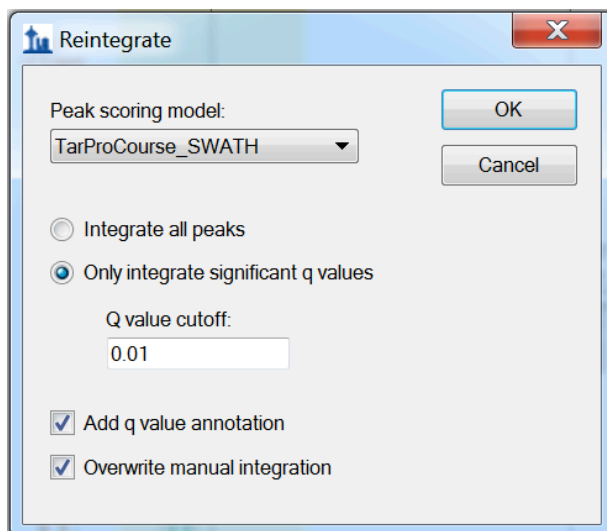
- Go to “View” → “Arrange Graphs” → “Grouped...”
- Group panes = 2, select: “Distribute graphs among groups”, Display: “Tiled”, Sort order: “Document” → “OK”
- Save your skyline document as *TarProCourse_Tut9_swath_extracted_data.sky*

2.3. mProphet

- Go to “Edit” → “Refine” → “Reintegrate”
- From the drop-down menu of the peak scoring model select: “Add..”
- Click “Train Model” and inspect the model score distributions
- Fill “TarProCourse_SWATH” as name
- The window should look like this:



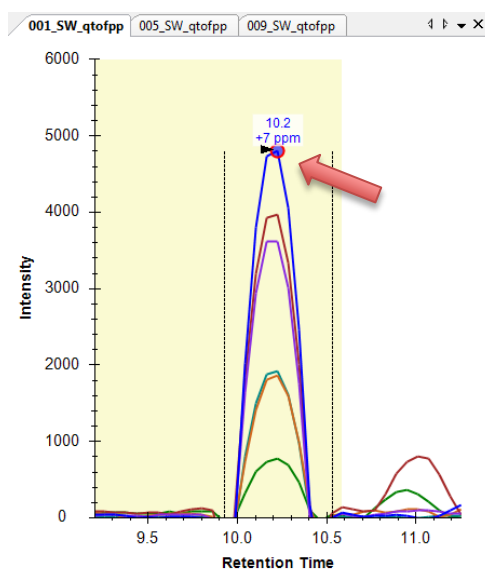
- Click “OK”
- Back in the “Reintegrate” window select following:



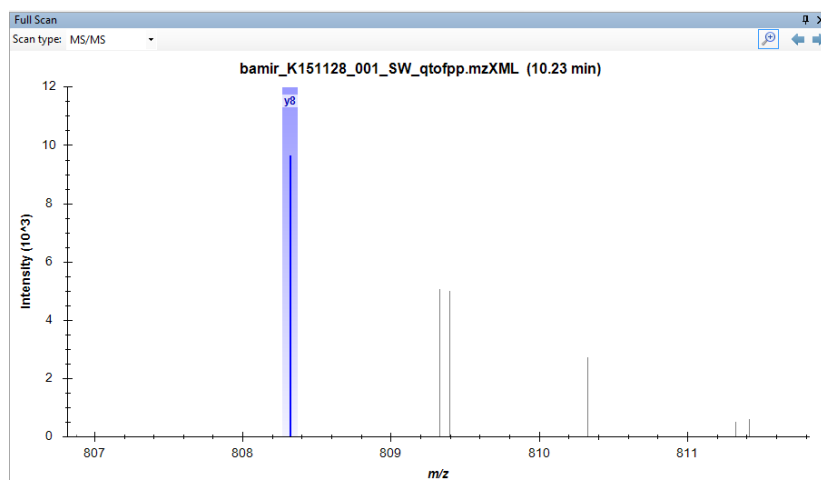
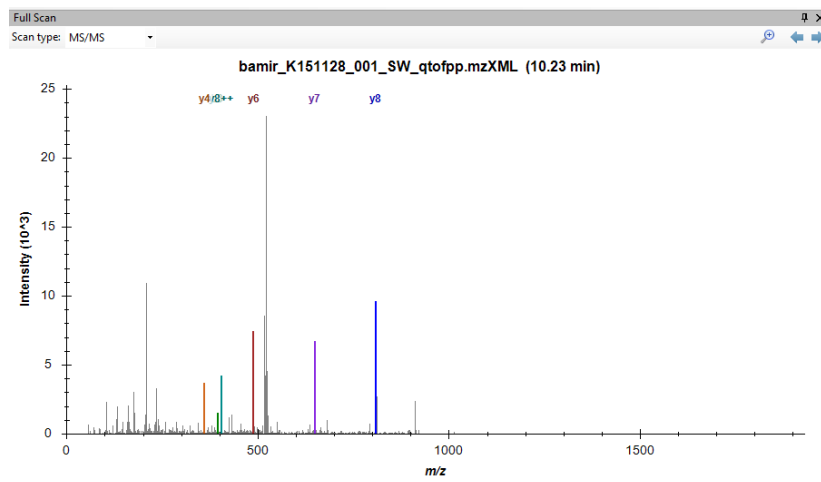
- Click “OK”
- Note that some target peptides now have a red dot in front of their sequence.
- Save your skyline document as *TarProCourse_Tut9_swath_mProphet.sky*

2.4. Inspect the data manually

- Now we will manually inspect some of the chromatography and underlying spectra
- Add the peak area and retention time views to your document if they are not present (View → Retention Times → Replicate Comparison and View → Peak Areas → Replicate Comparison)
- Search for the protein O33268 (you can press *ctrl-F* and type the protein accession to find this quickly)
- Examine the pattern of peak areas for the 3 peptides belonging to this protein between the wild type and mutant conditions. Is this protein differentially expressed?
- Hover the mouse over one of the extracted ion chromatograms from the peptide sequence SFCCGAGGAR until a small circle appears overlaid on the XIC. Click on the circle near the apex of the peak. A graph showing the full scan SWATH MS2 spectrum should appear.



- A graph showing the full scan SWATH MS2 spectrum for this retention time should appear. Click on the magnifying glass symbol to zoom in the fragment ion you selected and again to zoom out to the full scan spectrum. The colored areas indicates the width of the extraction window for the XICs. You can use the left and right scroll arrows to browse through the preceding or following scans.



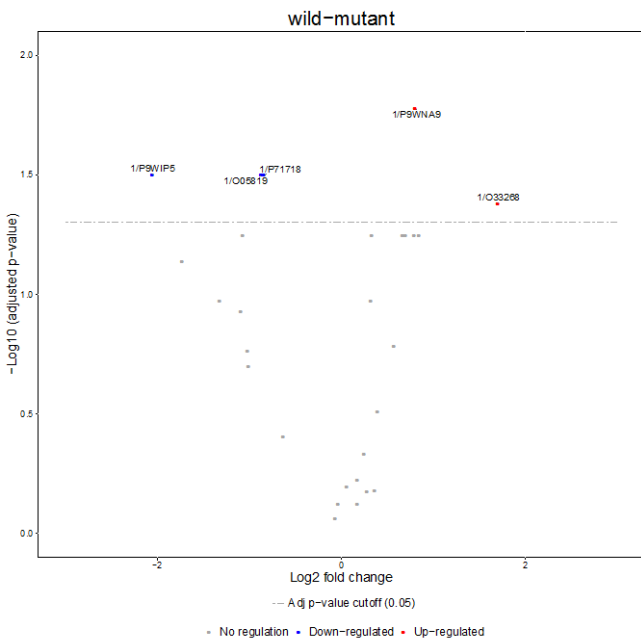
- Explore the data further manually (including some decoy assays).

2.5. MSstats

- Go to “Tools” → “MSstats” → “QC”
- Choose “None” as normalization method and leave the rest as default

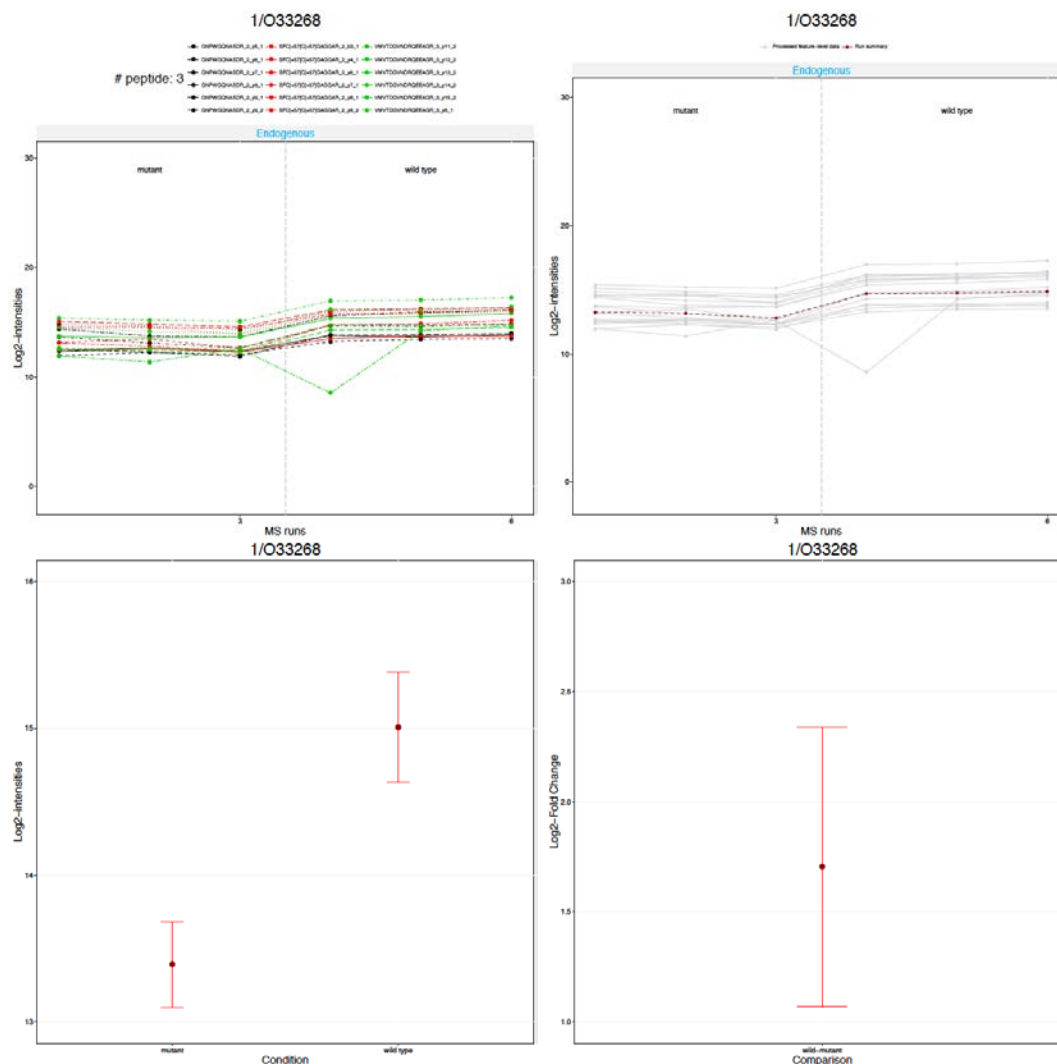
NOTE! We selected a very small library here without any spiked in peptide standards. Thus, normalization would remove the expected biological signals. In general we recommend you to use standard reference peptides or have large enough assay libraries where you expect the majority of proteins to have a constant quantity across all conditions.

- Click “OK”
- After “QC” is finished run “Group Comparison” and “Design Sample Size” applying the same parameters.
- You should get following output files:
 - 3 x MSstats log files for all 3 MSstats processing steps
 - 3 x dataProcessedData.csv
 - TestingResult.csv
 - SampleSizeCalculation.csv
 - QCPlot.pdf
 - ProfilePlot.pdf & ProfilePlot_wSummarization.pdf
 - ConditionPlot.pdf
 - ComparisonPlot.pdf
 - VolcanoPlot.pdf
 - SampleSizePlot.pdf
- Inspect the output files and compare, for example, how the information in the TestingResult.csv is represented in the volcano plot.



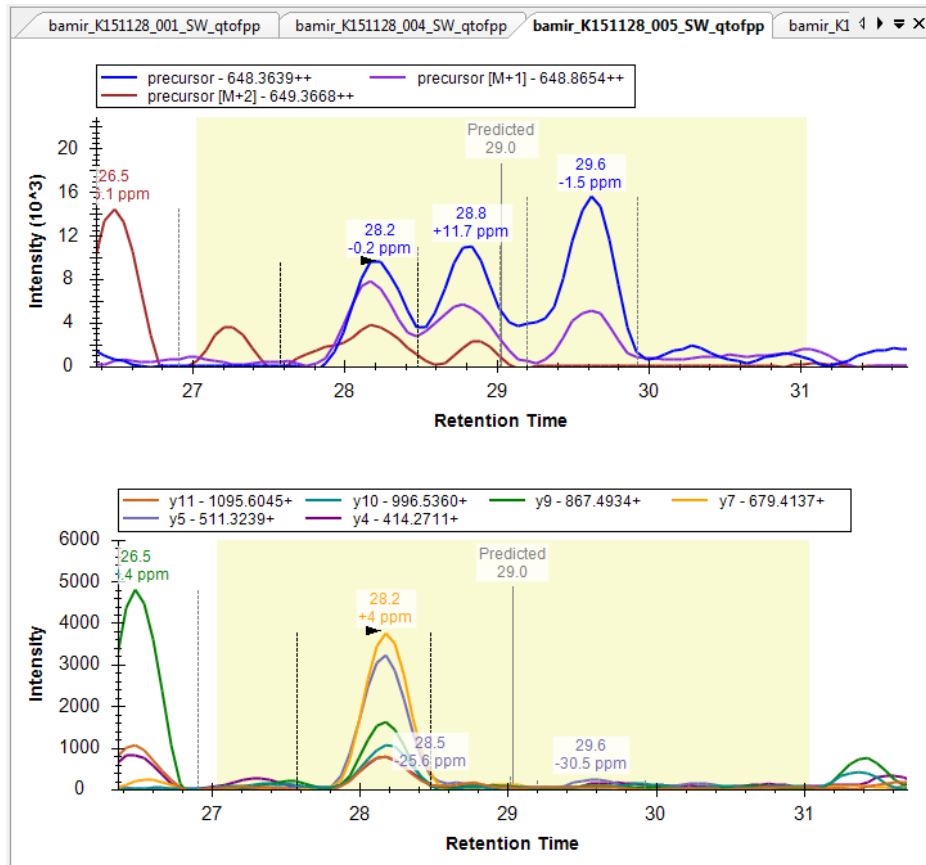
	A	B	C	D	E	F	G	H	I
	Protein	Label	log2FC	SE	Tvalue	DF	pvalue	adj.pvalue	
1	1/16Y0L1	wild-mutant	0.340006722	0.043020999	7.903273483	2	0.015635311	0.056502613	
2	1/O05819	wild-mutant	-0.869753664	0.05868556	-14.82057369	2	0.004521853	0.031652972	
3	1/O33268	wild-mutant	1.70356492	0.147633536	11.53914599	2	0.00742666	0.041589297	
4	1/O53638	wild-mutant	NA	NA	NA	NA	NA	NA	
5	1/O53692	wild-mutant	0.36805256	0.542578401	0.678339866	2	0.567519041	0.662105548	
6	1/O53695	wild-mutant	0.401457392	0.219345074	1.830254879	2	0.208699457	0.307557095	
7	1/P71716	wild-mutant	NA	NA	NA	NA	NA	NA	
8	1/P71718	wild-mutant	-0.839151867	0.047050786	-17.83502339	2	0.003129036	0.031652972	
9	1/P96239	wild-mutant	0.295252288	0.351451354	0.840094325	0	NA	NA	
10	1/P9WFX1	wild-mutant	-1.725715518	0.312588002	-5.520734977	2	0.031278766	0.072983787	
11	1/P9WHW3	wild-mutant	0.175780473	0.197473679	0.890146338	2	0.467308623	0.59475643	
12	1/P9WI43	wild-mutant	-0.03208297	0.069829727	-0.459445738	1	0.725819844	0.752702061	
13	1/P9WIP5	wild-mutant	-2.051673687	0.12506141	-16.40532988	2	0.003695028	0.031652972	
14	1/P9WIC7	wild-mutant	0.176828111	0.434211161	0.407239903	2	0.723282449	0.752702061	
15	1/P9WIE5	wild-mutant	0.698469601	0.085067501	8.210769031	2	0.014511035	0.056502613	
16	1/P9WIS9	wild-mutant	0.28470608	0.453489182	0.627812285	2	0.594253961	0.665564436	
17	1/P9WIV3	wild-mutant	0.854484116	0.122817001	6.957376471	2	0.020040069	0.056502613	
18	1/P9WK15	wild-mutant	-0.998941259	0.204499482	-4.884810704	1	0.128550243	0.199967045	
19	1/P9WKF7	wild-mutant	-0.057347535	0.060311208	-0.950860324	0	NA	NA	
20	1/P9WKZ5	wild-mutant	0.254555838	0.209169344	1.216984446	2	0.347726363	0.463635151	
21	1/P9WMM9	wild-mutant	0.327157967	0.077771591	4.206651318	2	0.052130756	0.106181349	
22	1/P9WNA9	wild-mutant	0.8081178	0.019740253	40.93756016	2	0.000596167	0.016692663	
23	1/P9WNE5	wild-mutant	0.068728439	0.089470001	0.768172999	2	0.522688794	0.636316793	
24	1/P9WPK3	wild-mutant	-0.064231198	0.326893372	-0.196489754	2	0.862382703	0.862382703	
25	1/P9WQ3	wild-mutant	0.802066455	0.117074943	6.850880623	2	0.020648636	0.056502613	
26	1/P9WNR3	wild-mutant	0.583347543	0.192403644	3.031894468	2	0.09374007	0.164045122	
27	1/P9WPI3	wild-mutant	0.673209798	0.097471584	6.906728807	2	0.020326125	0.056502613	
28	1/P9WPP5	wild-mutant	0.4696493	0.579514357	0.810418749	0	NA	NA	
29	1/P9WQ41	wild-mutant	-0.629456791	0.429346883	-1.466079798	2	0.280276832	0.392387565	
30	1/P9WQ63	wild-mutant	-1.066727614	0.16163315	-6.599683387	2	0.022197455	0.056502613	
31	1/P9WQF1	wild-mutant	-1.319199646	0.110269924	-11.96336772	1	0.053090675	0.106181349	
32	1/P9WQF9	wild-mutant	0.381226983	0.127490497	2.990238424	0	NA	NA	
33	1/P9WQJ7	wild-mutant	-1.015902652	0.356686557	-2.848166355	2	0.104334773	0.171845508	
34	1/P9WQJ9	wild-mutant	-1.080486251	0.284713702	-3.794992106	2	0.062949633	0.117505981	

- Check the results for the protein O33268 which we inspected manually earlier. What does the statistical analysis indicate?



2.6. Comparing MS1 and MS2 data

- Until now we have only been looking at extracted ion chromatograms from MS2 data (i.e. fragment ions). Now we will compare the signals from MS1 XICs and MS2 XICs
- Add the precursor ion XICs by clicking 'Edit' → 'Refine' → 'Advanced' and clicking on 'Auto-select all: Transitions'. This will add in the MS1 XICs from the first 3 isotopologues
- The MS1 XICs from the precursor ions will now be overlaid with the SWATH MS2 fragment ion XICs from the fragment ions. In general, the precursor ions have a much higher absolute intensity than the fragment ions. As such, it is usually easier to view these in separate graphs (View → Transitions → Split Graph)
- Browse through some peptides and compare the signals coming from the MS1 XICs and the SWATH MS2 XICs. Is there a difference in the selectivity between the MS1 and SWATH MS2 data
- Search (using *ctrl-F*) for the peptide SIVESTPAPPGIK and examine the difference between the MS1 and MS2 data.



We would like to thank SystemsX for supporting the Zurich Targeted Proteomics Course 2016.



SystemsX.ch
The Swiss Initiative in Systems Biology