Tutorial 6: Manual SRM data analysis

After the generation of a transition list and its measurement by SRM, the next step is quality control and data analysis of your SRM runs. Skyline offers very useful graphical views that allow a fast and straightforward peak intensity and retention time comparison over many SRM runs.

Tip! If you plan a large SRM measurement including lots of samples, it is well worth to invest time upfront on defining the optimal transition list for your samples in some refinement runs. This helps to avoid repeated measurement of transitions which are of bad quality or not informative.

Tip! Import each data file into Skyline directly after the measurement to monitor data quality, sensitivity and reproducibility. This allows taking immediate action in case instrument problems occur.

Tip! Select a few positive control peptides from the background proteome, which are monitored in each run and allow controlling sample quality and instrument performance over months and years. Ideally, positive control peptides do not change in abundance over the samples, i.e. they should come from housekeeping proteins. This allows normalisation and thus accounting for variability in sample preparation etc. Further, it is useful if these peptides cover a wide abundance range (high, medium, and low abundant peptides).

Label-based SRM data analysis

Importing raw data into Skyline

- Open the Skyline document *TarProCourse_Tut5_iRT_scheduled.sky* from the folder *Tutorial-5_Scheduling.*
- Save it as *TarProCourse_Tut6_label_raw.sky* to the folder *Tutorial-6_ManualAnalysis.*
- Remove the retention time calibration run: "Edit" → "Manage results" → "Remove" *olgas_S130701_001_A1_RTcal.*
- Import the target runs: "File" → "Import" → "Results" → "Add single injection replicate per files" → OK → Select all nine .wiff files in folder *Tutorial-6_ManualAnalysis* → "Open".
    - 2 of these wiff files contain several runs:
        - olgas_S130701_002_A1.wiff → Select A1 (2) only.
        - olgas_S130701_003_B1.wiff → select B1 (3) only.
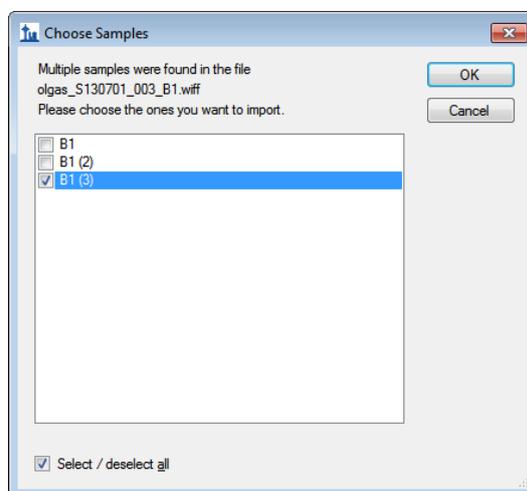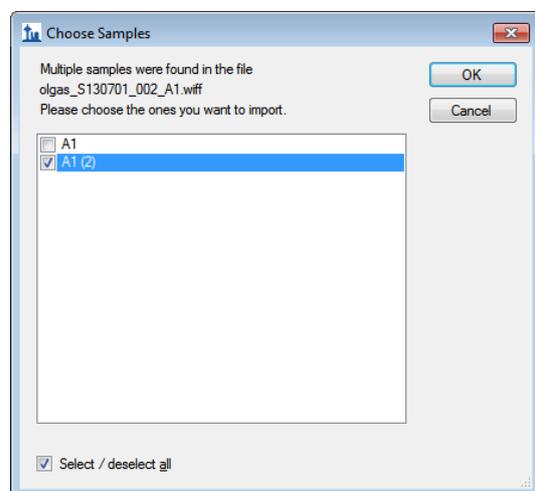
- To rename the runs, go to "Edit" → "Manage results" → double-click onto file name and replace with new one:
    - A1 (2)→ 0h_rep1
    - B1 (3) → 6h_rep1
    - olgas_S130701_004_C1 → 48h_rep1
    - olgas_S130701_005_A2 → 0h_rep2
    - olgas_S130701_006_B2 → 6h_rep2
    - olgas_S130701_007_C2 → 48h_rep2
    - olgas_S130701_008_A4 → 0h_rep3
    - olgas_S130701_009_B4 → 6h_rep3
    - olgas_S130701_010_C4 → 48h_rep3

- Go to "Settings" and activate "Integrate all" to ensure that always all transitions are integrated.
- As some transitions from the Skyline document were not actively measured in the SRM measurements (issue occurring due to software version variances in selection of the most-intense transition from the spectral library) we quickly want to get rid of the missing values. For this go to "Edit" → "Refine" → "Remove missing Results".
  In the Skyline file 442 transitions will remain.
- Activate the peak area and retention time view (if not already there):
    - "View" → "Retention times" → "Replicate comparison"
    - "View" → "Peak areas" → "Replicate comparison"
    - **TIP!** Once these windows are open, you can right-click on them to choose the type of graph.
- Skyline offers you to arrange all open windows in your preferred way.
    - Position the "Peak Areas" and "Retention Times" windows below the SRM chromatograms.
    - Your
    - Place the "Library Match" window below the "Targets" (transition-tree) window.
    - Go to "View" → "Arrange graphs" → "Tiled", now you see all the imported SRM runs at once.
    - There are different options to arrange your graphs and everyone should choose whatever is most convenient for him/her. For now, arrange the three biological replicates in three windows and sort the different time points for each replicate in tabs. In order to do so go to:
      "View" → "Arrange graphs" → "Grouped"
        - Group panes: "3"
        - Separate graphs into groups
        - Sort order: "Document"
        - "OK"
            -
    - Your Skyline document should now look like this:
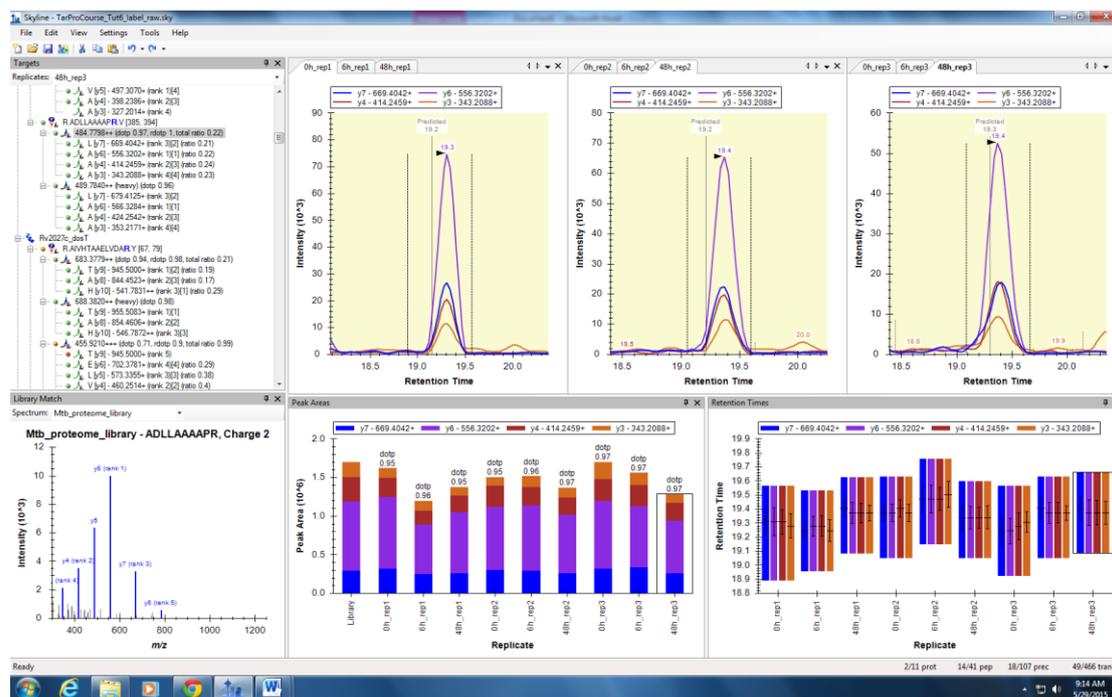
- Save the file

Manually explore, correct and refine the picked SRM chromatograms

- First check the reproducibility of the MS performance by inspecting the signals of the iRT peptides over all 9 runs. After going through the iRT peptides, continue with the 30 target peptides. We recommend to in a first round just refine the peak picking and in a second round refine the transitions for quantification:

  Peak picking (identification)
    o Are all peptides picked correctly by Skyline?
    o Does the iRT-predicted retention time fit the actually measured peptide retention time?
    o Does the picked peak match in relative transition intensities to the library spectrum (good dot-product)?
    o Consider also the information from the heavy channel!
    o Are the peak boundaries set appropriately? In the peptide retention time replicate view all peptide bars should have approximately the same height. If certain precursors were not picked correctly, drag the area boundaries to what you think is the right peak.
    The peak boundaries for all precursors belonging to one peptide have to be the same (Skyline should automatically keep precursor boundaries the same, but always check).
    For some peptides it happens that they are not detectable in the first time points and only come up during later time points. In this case you can integrate the background as long as you are sure, that you are in the correct region.
    Tip! When you right-click onto the chromatograms, you have several options to customise visualisation. Try them all to see what happens. We often switch between different zooming (None / Best Peak) and transformation (Savitzky-Golay Smoothing).
    → After you are sure that all peaks are picked correctly save the file as *TarProCourse_Tut6_label_picked.sky*.

Transition refinement (quantification)
- o Are all transitions of good quality and reproducible over the samples? The relative transition intensity has to be constant over all runs. To visualise this, right-click on the "Peak Areas" window and select "Normalized to total".
- o If certain transitions/precursors/peptides are of low quality (low intense, not co-eluting with the other transitions, shouldered, etc.) or irreproducible over runs, remove them from the document by deleting them from the "Targets" (transition tree) window.
  Tip! You can bring back deleted transitions/precursors/peptides by right-clicking on the respective parent item → "Pick children".
→ Save the refined file as *TarProCourse_Tut6_label_refined.sky*.

Export a custom-defined report

- • Export all quantitative data to a .csv file for further processing.
  - o "File" → "Export" → "Report"
  - o The "Report" tab offers you a range of predefined report formats. Here we will generate our own report format: "Edit list" → "Add..."
  - o An "Edit Report" window opens.
  - o **Note!** For detailed information about all options see the Skyline tutorial "Skyline Custom Reports and Results Grids" on the Skyline website.
  - o Name the report "TarProCourse_quant_TotalArea". From the selection tree on the left, select the following items:
    - ▪ Proteinname
    - ▪ PeptideModifiedSequence
    - ▪ Ratio to Standard
    - ▪ Add a check mark to "Pivot Replicate Name" in the bottom of the window.
  - o Your Edit Report tab should now look like this:
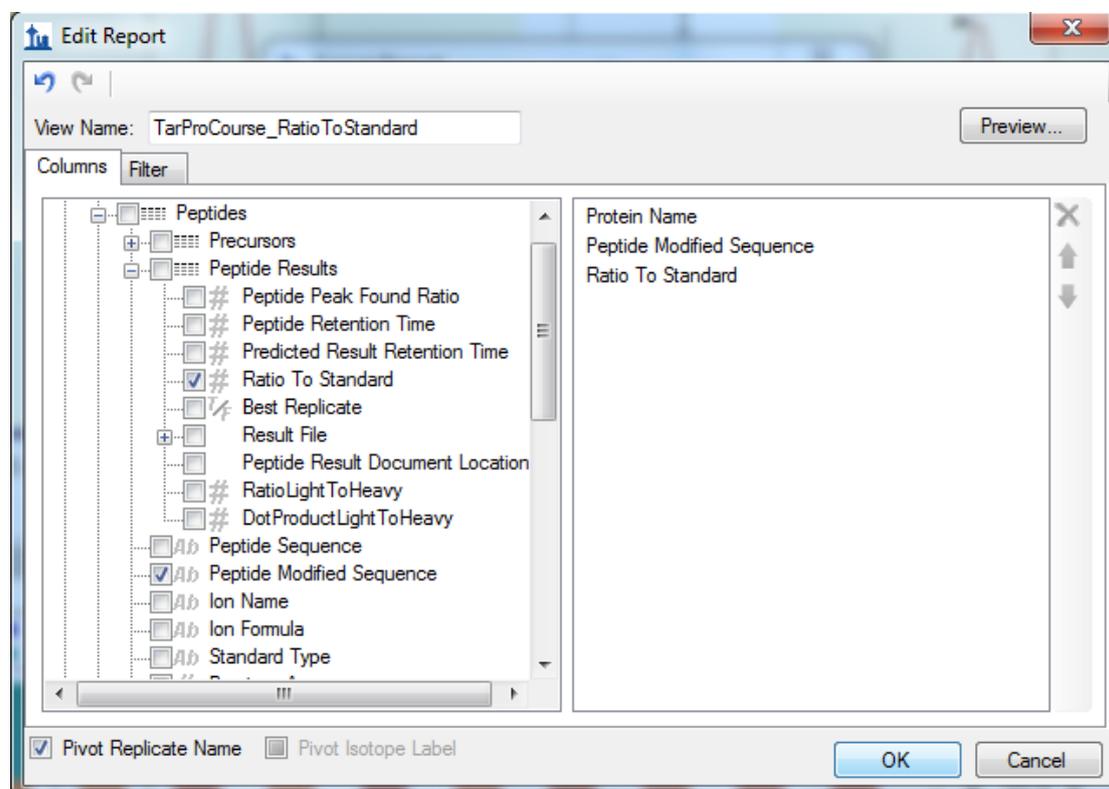
- o Click the "Preview" button to see the table that will be exported.
- o If you are happy with the report format, save it by clicking "OK".

- Export the data using this newly generated report format and save it as *TarProCourse_Tut6_label_refined.csv*.
- Save the Skyline file.

Exercises

1. What are advantages and disadvantages of label-free and label-based SRM data analysis?

2. How important is the careful placement of the peak boundaries in label-free compared to label-based data analysis?

3. Why should the ratio between a spike-in and endogenous peptide be around 1 for accurate protein quantification?

4. Data analysis in Excel (only if time allows):

   Statistics/Excel reminder

   - **Standard deviation**: The standard deviation is a measure of how widely values are dispersed from the average value (the mean).
     → Standard deviation in Excel: STDEV(number1, number2, …)
   - **Coefficient of variance (CV)**: Normalise (divide) standard deviation by average and multiply by 100 to obtain the result in %.

   Label-based SRM data analysis

   1. Open the file *TarProCourse_Tut6_label_refined.csv*.
   2. Remove all iRT peptides and reorder the columns such that the replicates are next to each other.
   3. Peptide level:
      a. Calculate the Ratio-To-Standard ratio of each of the 9 samples to the averaged Ratio-To-Standard of time point 0h for each peptide.
      b. Calculate the average, the standard deviation, and the CV of the Ration-To-Standard ratios over the 3 biological replicates per time point.
   4. Protein level:
      a. Average the Ratio-To-Standard ratio of all precursors per protein for each of the 9 samples.
      b. Calculate the average protein ratio, the standard deviation, and the CV over the three biological replicates.
   5. Plot the ratios from 4b and change the y-axis to log2 scale for better visualisation.
      Note! Log2 scale is advantageous because the scale is compressed for big changes and enlarged for small changes and because the ratios for down-regulation are symmetrical to the ones for up-regulation, whereas in non-log scale the ratios for down-regulation are between 0 and 1, whereas the ratios for up-regulation are between 1 and infinite.
   6. What is the average peptide-level CV over the whole dataset?

SystemsX.ch
The Swiss Initiative in Systems Biology