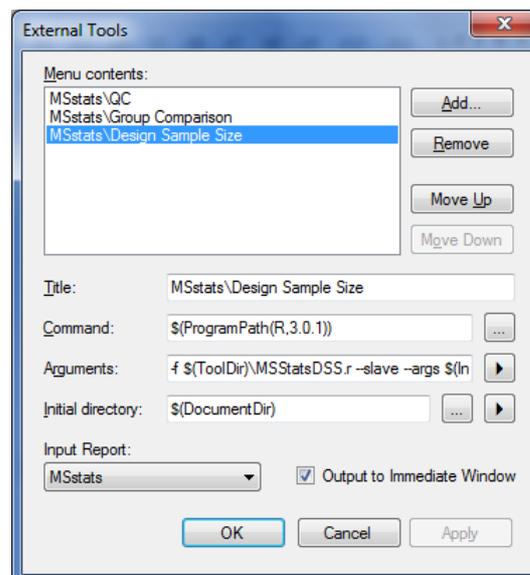


## Tutorial 8: MSstats

### 1. Installing MSstats as an External Tool

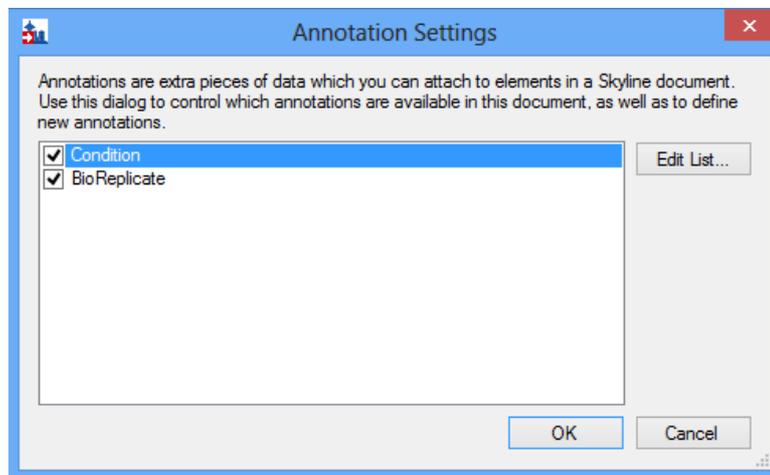
- The Skyline external tools framework supports the direct integration of statistical analysis tools that process the data in a Skyline document.
- To be able to run MSstats, Skyline will install R3.0.2 on your laptop. We experienced issues when older versions of R were installed and we therefore recommend to uninstall previous R versions through “Control panel”. Please also delete the contents of the folder: C:/Users/xxx/Documents/R/win-library/ (where xxx is user user name).
- To add MSstats as an external tool, perform the following steps:
  - On the “Tools” menu, click “Tools Store...”.
  - On the left, navigate to the “MSstats” tool to see the installation status and a description.
  - Click “Install”. Skyline will now guide you through the process of installing MSstats as an external tool. This process may require you to install the R statistical programming environment, as well as a number of R packages which are used by MSstats to perform its statistical analysis.
    - On the newly opened R Installer form, click “Install”.
    - If you do not have the specified version of R installed on your machine, wait for R to download.
    - When the download is completed, you will be asked to install R onto your computer.
    - Complete the R installation process by working through the R installer wizard.
    - Once the R installation is completed, Skyline will download and install the necessary R packages for MSstats. When prompted to run the “SkylineProcessRunner”, click the “Yes” button to install the packages.
  - Once the installation is complete, the MSstats tools will appear in the External Tools form (there might be already other external tools in the list):



- Click the “OK” button on the External Tools form.

## 2. Annotating samples with group information

- Open the Skyline file *SRMcourse\_20140212\_label\_refined.sky* in the folder *Tutorial-6\_ManualAnalysis*.
- Before we can run MSstats, we need to annotate which samples are replicates. Skyline allows you to associate additional information with the runs in the document by defining custom annotations. To view the Annotation Settings form, perform the following steps:
  - On the “Settings” menu, click “Annotations...” → “Edit List...” → Select “Condition” and click “Edit...”:
    - Name: Condition
      - Type: Value List
      - Values: “00h”, “06h”, “48h”, each on a new line (without “ ”)
      - Applies To: Replicates
      - Click twice “OK” to confirm.
  - After confirming the new annotation type, activate the 2 required Annotations for MSstats by checking their checkboxes: Condition and BioReplicate (with default settings).
  - Click the “OK” button.



- Editing annotation values in Skyline is done using the Results Grid.
  - To bring up the Results Grid go to the “View menu and click “Results Grid”. The Results Grid will show you chromatogram peak areas and other measured results for the currently selected peptide, or transition.
  - Annotate “Condition”, “BioReplicate” as shown in the table below.
 

**Caution!** This table view will only be visible when a peptide is selected in the target window. If a protein, a precursor or a transition is selected the columns will change accordingly.

**Caution!** If you see the little pencil icon at the left bottom, move the cursor to any other cell to save all information.

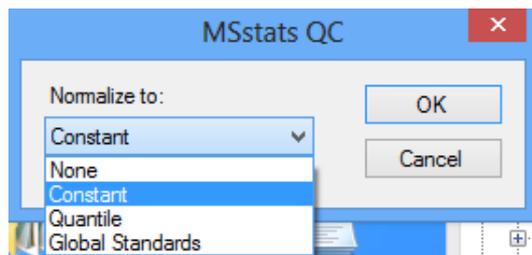
Replicate	Condition	BioReplicate
0h_rep1	00h	1
6h_rep1	06h	1
48h_rep1	48h	1
0h_rep2	00h	2
6h_rep2	06h	2
48h_rep2	48h	2
0h_rep3	00h	3
6h_rep3	06h	3
48h_rep3	48h	3

Replicate	Condition	BioReplicate
0h_rep1	00h	1
6h_rep1	06h	1
48h_rep1	48h	1
0h_rep2	00h	2
6h_rep2	06h	2
48h_rep2	48h	2
0h_rep3	00h	3
6h_rep3	06h	3
48h_rep3	48h	3

- Close the Results Grid.
- Save the Skyline file as *SRMcourse\_20140213\_label\_MSstats.sky* in the folder *Tutorial-8\_MSstats*.

### 3. Running MSstats

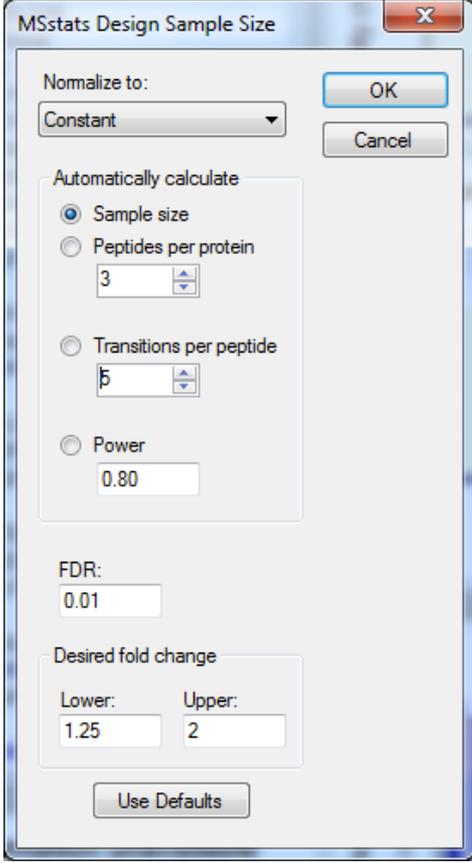
- MSstats is composed of three individual tools:
  - **Quality Control (QC)**: Provides quality control statistics for MS runs that become quantitative data for next step analysis. Logarithm transformation with base 2 and then normalization to remove systematic bias between MS runs is applied.
  - **Group Comparison**: Tests for significant changes in protein abundance across conditions based on linear mixed-effects models.
  - **Design Sample Size**: Calculates sample size for future experiments using an intensity-based linear model.
  - **Tip!** For more detailed information about functionality and options, visit [msstats.org](http://msstats.org).
- To run a QC analysis, perform the following steps:
  - On the “Tools” menu, choose “MSstats” and click “QC”.
  - Choose ‘Constant’ for normalization.



- Skyline will begin to export the MSstats Input report. Once the report is exported, Skyline will output the results of the tool run in the Immediate Window docked at the bottom of the main Skyline Window.
- When the tool is completed (“Finished.” In the Immediate Window), navigate to the directory containing the *SRMcourse\_20140213\_label\_MSstats.sky* file. This directory should now contain the outputs of the dataProcess: dataProcessedData.csv, ProfilePlot.pdf, QCPlot.pdf, ConditionPlot.pdf. There are also 2 txt files containing information about the software/package versions: MSstats\_process.txt and sessionInfo.txt.
- To run a Group Comparison analysis, perform the following steps:
  - On the “Tools” menu, choose “MSstats” and click “Group Comparison”.
  - Choose ‘Constant’ in Normalize to.
  - In the Name of comparison textbox enter “6h-0h”.
  - Click 00h in Control group.
  - Click the “06h” group to compare against.
  - Keep the default settings.
    - Label data: whether label-free or with isotope labelled reference.
    - Assume equal variance: whether we can assume equal variance among transitions.
    - Include interference transitions: whether interference transitions will be included for fitting the model or not.
    - Scope of biological replicate: the conclusion should be extended beyond the subjects selected for the study, so at the level of the underlying biological populations (“expanded”) or the conclusion

- should be restricted to the subjects selected for the study (“restricted”).
    - Scope of technical replicate: the conclusion should be extended beyond the mass spectrometry runs performed in the study, so at the level of the underlying populations of possible replicates of mass spectrometry runs (“expanded”) or the conclusion should be restricted to the performed mass spectrometry runs (“restricted”).
  - Click the “OK” button of the MSstats Group Comparison form.

- The Group Comparison tool will now perform its analysis. The output from the tool is shown in the Immediate Window. When the group comparison run is completed, switch back to the directory containing the *SRMcourse\_20140213\_label\_MSstats.sky* file. The directory should now contain VolcanoPlot, ComparisonPlot PDF files, testing result CSV file, and an MSstats\_process-1.txt file generated by the tool.
- To run a sample size calculation analysis, perform the following steps:
  - On the Tools menu, choose MSstats and click Design Sample Size.
  - Choose ‘Constant’ in Normalize to.
  - For this tutorial you will leave the options in their default settings.
    - The default setting will calculate sample size when there is 3 peptides per protein, 5 transitions per peptide in order to get power=0.8 with FDR=0.01 between 1.25 and 2 fold change.
    - One of “Sample size”, “Peptides per protein”, “Transitions per peptide”, and “Power” need to be selected to calculation. Other three values should be provided.
  - Click the OK button of the MSstats Design Sample Size form.



The image shows a dialog box titled "MSstats Design Sample Size". It has a standard Windows-style title bar with a close button (X) in the top right corner. The dialog is divided into several sections:

- Normalize to:** A dropdown menu set to "Constant". To the right are "OK" and "Cancel" buttons.
- Automatically calculate:** A group box containing four radio button options:
  - Sample size
  - Peptides per protein: A spin box with the value "3".
  - Transitions per peptide: A spin box with the value "5".
  - Power: A text box with the value "0.80".
- FDR:** A text box with the value "0.01".
- Desired fold change:** A group box with two text boxes:
  - Lower: "1.25"
  - Upper: "2"
- Use Defaults:** A button at the bottom of the dialog.

- The Design Sample Size tool will now perform its analysis. The output from the tool run is shown in the Immediate Window. When the tool is completed, switch to the directory containing the *SRMcourse\_20140213\_label\_MSstats.sky* file. The directory should now contain a *SampleSizePlot* PDF file, *SampleSizeCalculation* CSV file, and a *MSstats\_process-2.txt* file generated by the tool.

## Exercises

1. Please try Quantile normalization in `dataProcess` and check how different the QC plots are.
2. Please try group comparison for 48h vs 6h.
3. How is the result of testing changed with expanded scope of biological replicate for 48h vs 6h?
4. Have a look at the `SampleSizePlot.pdf`. What is the minimum number of replicates needed, if we would like to confidently (with  $\text{power}=0.8$ ) quantify a fold-change of 1.4?
5. Please calculate the power when there are measurements with 3 subjects, 2 peptides per protein, 3 transitions per peptide for  $\text{FDR}=0.05$ .

We would like to thank Prime-XS and SystemsX for supporting the Zurich SRM Course 2014.

