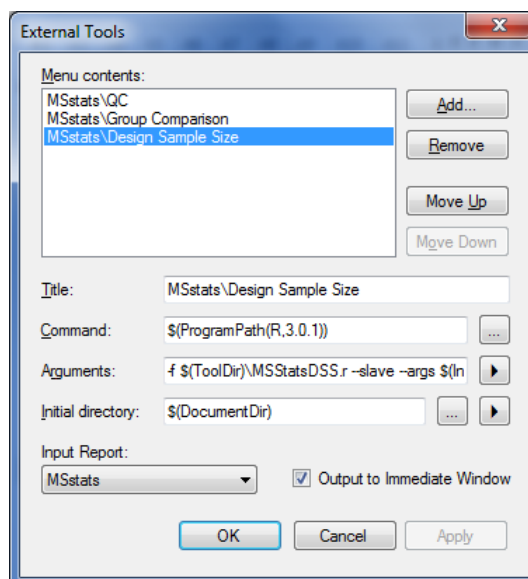


MSstats as an External Tool

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 add MSstats as an external tool, perform the following steps:
 - On the “Tools” menu, click “External Tools” to bring up the External Tools form.
 - Click the “Add...” button, and select “From Tools Store...” from the menu that appears.
 - 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.

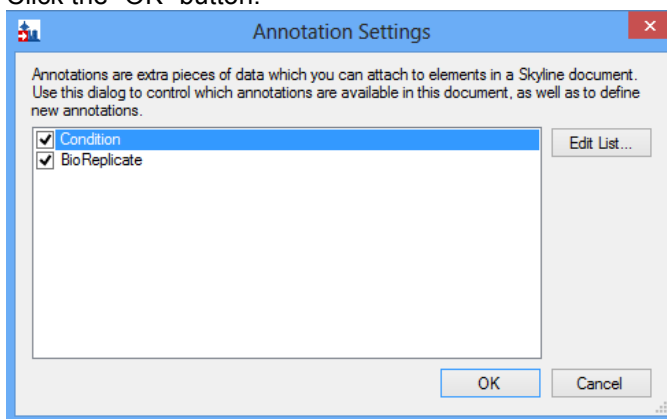
!! Troubleshooting

If you have a problem installing required packages from **Tool store** in Skyline, there are two possible solutions:

1. Uninstall previous version of R through **Control panel** on window system. Then also delete R folder under 'C:/Users/XXX/Documents/' (where XXX is your user name) or under 'C:/Program files/'. Re-install R and MSstats.
2. If you still have a problem, try to pause your **antivirus** program while installing the packages because antivirus program interrupt to install required R packages. Then install MSstats again.

2. Annotating samples with group information

- Example data : rat plasma for heart failure
 - 7 control samples and 7 disease samples
 - 3 technical replicates
 - Total 42 injections (runs)
 - Label-free SRM
- Open the zipped Skyline file *Rat_plasma.sky* in the folder *Tutorials/GroupedStudies/Heart Failure* of Flash drive.
- 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: "Disease", "Healthy", 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 (default).
 - Click the "OK" button.



- Editing annotation values in Skyline is done using the Results Grid. To bring up the Results Grid do the following:
 - On the "View" menu, 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. (Please check whether there is any value missing or wrong) (**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) :

Results Grid

Views ▾ | 1 of 42 | Export... | Find: ▾


	Replicate	Condition	BioReplicate
▶	D_102_REP1	Disease ▾	1
	D_102_REP2	Disease ▾	1
	D_102_REP3	Disease ▾	1
	D_103_REP1	Disease ▾	2
	D_103_REP2	Disease ▾	2
	D_103_REP3	Disease ▾	2
	D_108_REP1	Disease ▾	3
	D_108_REP2	Disease ▾	3
	D_108_REP3	Disease ▾	3
	D_138_REP1	Disease ▾	4
	D_138_REP2	Disease ▾	4
	D_138_REP3	Disease ▾	4
	D_154_REP1	Disease ▾	5
	D_154_REP2	Disease ▾	5
	D_154_REP3	Disease ▾	5
	D_172_REP1	Disease ▾	6
	D_172_REP2	Disease ▾	6
	D_172_REP3	Disease ▾	6
	D_196_REP1	Disease ▾	7
	D_196_REP2	Disease ▾	7
	D_196_REP3	Disease ▾	7

Results Grid

Views ▾ | 1 of 42 | Export... | Find: ▾

	Replicate	Condition	BioReplicate
	H_146_REP1	Healthy ▾	8
	H_146_REP2	Healthy ▾	8
	H_146_REP3	Healthy ▾	8
	H_147_REP1	Healthy ▾	9
	H_147_REP2	Healthy ▾	9
	H_147_REP3	Healthy ▾	9
	H_148_REP1	Healthy ▾	10
	H_148_REP2	Healthy ▾	10
	H_148_REP3	Healthy ▾	10
	H_159_REP1	Healthy ▾	11
	H_159_REP2	Healthy ▾	11
	H_159_REP3	Healthy ▾	11
	H_160_REP1	Healthy ▾	12
	H_160_REP2	Healthy ▾	12
	H_160_REP3	Healthy ▾	12
	H_161_REP1	Healthy ▾	13
	H_161_REP2	Healthy ▾	13
	H_161_REP3	Healthy ▾	13
	H_162_REP1	Healthy ▾	14
	H_162_REP2	Healthy ▾	14
	H_162_REP3	Healthy ▾	14

Caution!!: If you can see 'pencil' at the left bottom, please move the cursor to any other cell. Then all information will be saved.

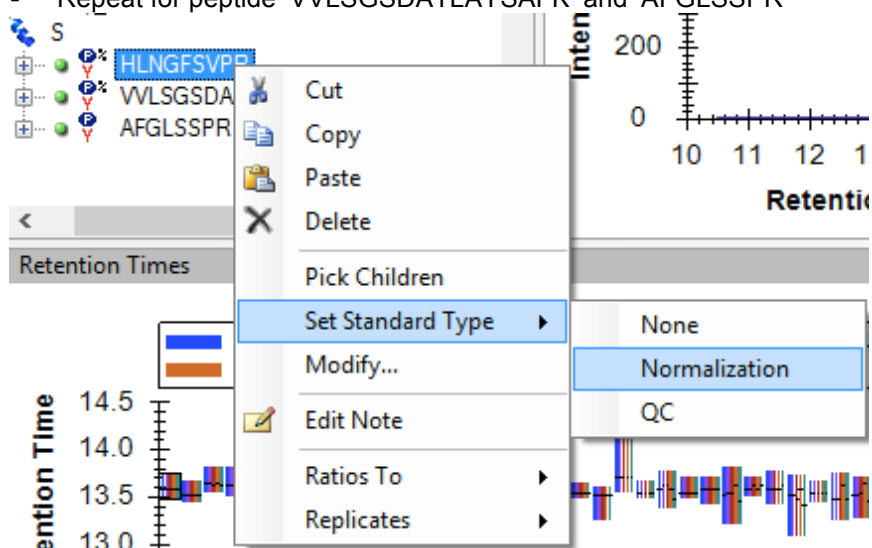
	G10	Disease ▾	79
	G11	Disease ▾	80
	G12	Disease ▾	81

	G10	Disease ▾	79
▶	G11	Disease ▾	80
	G12	Disease ▾	81

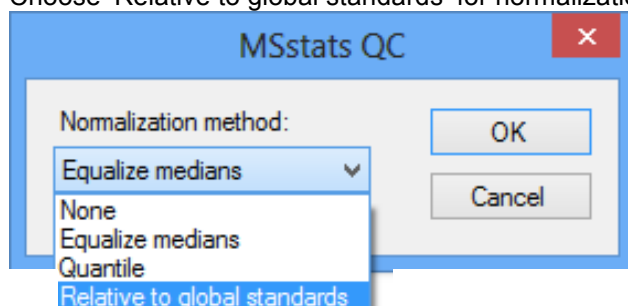
- Close the Results Grid.
- Save the Skyline file as *Rat_plasma.sky* in the folder.

3. Using 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.
- To run a QC analysis, perform the following steps:
 - For normalization with global standards, choose global standard proteins or peptides.
 - Click peptide 'HLNGFSVPR'
 - Right click and select 'Set Standard Type'
 - Choose 'Normalization'
 - Then '%' is shown on left side of peptide names
 - Repeat for peptide 'VLSGSDATLAYSAFK' and 'AFGLSSPR'

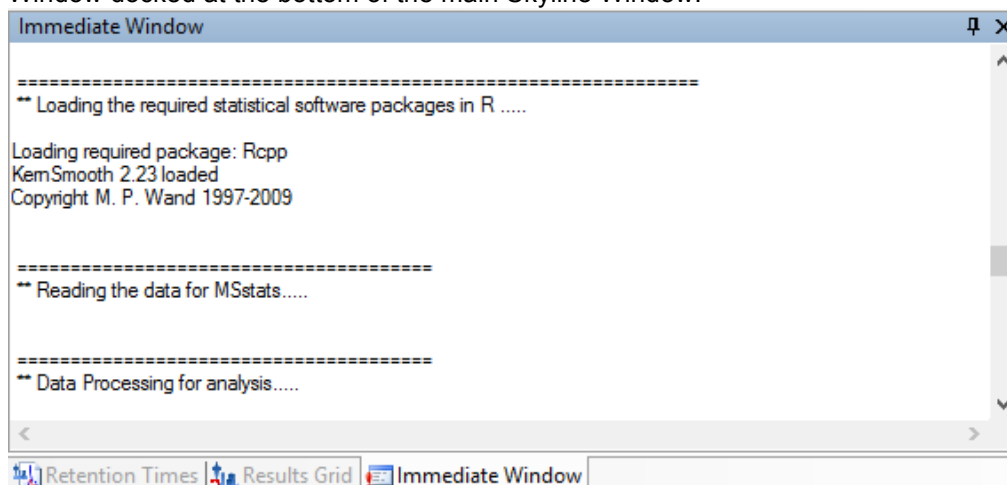


- On the "Tools" menu, choose "MSstats" and click "QC".
- Choose 'Relative to global standards' for normalization method.



- Keep the default settings.
 - Allow missing peaks: whether adding incomplete rows in the input or not.

- Skyline will begin exporting 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.



```
=====  
** Loading the required statistical software packages in R .....  
  
Loading required package: Rcpp  
KernSmooth 2.23 loaded  
Copyright M. P. Wand 1997-2009  
  
=====  
** Reading the data for MSstats.....  
  
=====  
** Data Processing for analysis.....  
  
< >
```

- When the tool is completed, navigate to the directory that has *Rat_plasma.sky* in the folder (*Tutorials/GroupedStudies/Heart Failure*). This directory should now contain process data with .csv file, ProfilePlot, QCPlot, ConditionPlot PDF files, and msstats.log generated by the QC tool.
- To run a Group Comparison analysis, perform the following steps:
 - On the “Tools” menu, choose “MSstats” and click “Group Comparison”.
 - Choose ‘Relative to global standards’ in Normalization method.
 - In the Name of comparison textbox enter “Disease-Healthy”.
 - Click ‘Healthy’ in Control group.
 - Choose settings as below.
 - Uncheck ‘Include reference standards’ : label-free experiment. If check, experiment with isotope labeled reference.
 - Check ‘Assume equal variance’ : whether we can assume equal variance among transitions.
 - Uncheck ‘Include interference transitions’ : whether interference transitions will be included for fitting the model or not.
 - Scope of biological replicate : Choose ‘Expanded’. The conclusion should be extended beyond the subjects selected for the study, so at the level of the underlying biological populations (“expanded”) or conclusion to the subjects selected to the study (“restricted”).
 - Scope of technical replicate : Choose ‘Expanded’. 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 conclusion to the performed mass spectrometry runs (“restricted”).
 - Click the “OK” button of the MSstats Group Comparison form.

The screenshot shows the 'MSstats Group Comparison' dialog box. The 'Name of comparison' field is set to 'Disease-Healthy'. The 'Normalization method' dropdown is set to 'Relative to global standards'. The 'Control group' dropdown is set to 'Healthy'. There are several checkboxes: 'Allow missing peaks' is unchecked, 'Include reference standards' is unchecked, 'Assume equal variance' is checked, and 'Include interference transitions' is unchecked. There are two radio button groups: 'Scope of biological replicate' with 'Expanded' selected, and 'Scope of technical replicate' with 'Expanded' selected. 'OK' and 'Cancel' buttons are located on the right side of the dialog.

- The Group Comparison tool will now perform its analysis. The output from the tool run is shown in the Immediate Window. When the group comparison run is completed, switch back to the directory that has *Rat_plasma.sky* file again. The directory should now contain VolcanoPlot, ComparisonPlot PDF files, testing result CSV file, and msstats.log 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 'Relative to global standards' in Normalization method.
 - For this tutorial you will leave the options in their default settings.
 - The default setting will calculate sample size when there is 2 peptides per protein, 3 transition per peptide in order to get power=0.8 with FDR=0.05 between 1.25 and 1.75 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 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 back to the directory that has *Rat_plasma.sky* file again. The directory should now contain a SampleSizePlot PDF file, SampleSizeCalculation CSV file, and msstats.log file generated by the tool.

Exercises

1. Please try Quantile normalization in dataProcess and check how different the QC plots are.
2. Have a look at the SampleSize.pdf. What is the minimum number of replicates needed, if we would like to confidently (with power=0.8) quantify a fold-change of 1.4?
3. Please calculate the power when there are measurements with 3 subjects, 2 peptides per protein, 3 transitions per peptide for FDR=0.05.