

MSstats: Statistical analysis of an SRM experiment with group comparisons

This tutorial will guide you through the installation of MSstats, and through using MSstats for a statistical analysis starting from a sample Skyline file. For more detailed examples and documentation, please visit msstats.org

Getting Started

To start this tutorial, download **MSstatsTutorial.zip** from this page:

<https://skyline.gs.washington.edu/labkey/skyts/home/software/Skyline/tools/toolDetails.view?name=MSstats>

Extract the files to a folder on your computer, e.g.:

C:\Users\ljae\Desktop

This will create a new folder:

C:\Users\ljae\Desktop\MSstatsTutorial

The zip file contains the following file:

Human_Plasma.zip: Contains a Skyline document containing sample data for MS stats.

You will use the Human Plasma dataset as an example. To open this dataset perform the following steps:

- Open **Skyline**.
- On the **File** menu, click **Open**.
- Change the **file type** in the **File name** field from 'Skyline Documents (*.sky)' to 'Shared Files (*.zip)'
- Navigate to the 'Human_plasma.zip' file in the folder you created and open it.

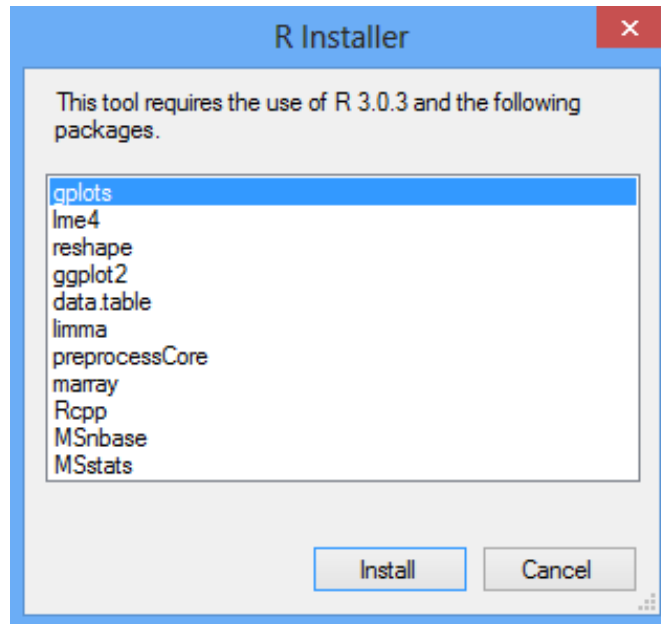
Installing MSstats as an External Tool

The Skyline external tools framework supports the direct integration of statistical analysis tools that process the data in a Skyline document.

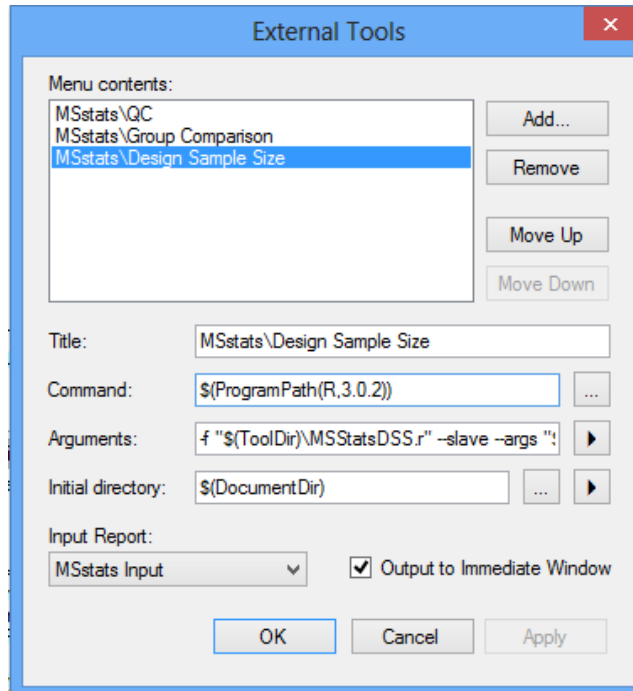
To add MSstats as an external tool, perform the following steps:

- On the **Tools** menu, click **Tool Store...** to bring up the **Install from Tool Store** form.
- On the left, navigate to the **MSstats** tool to see the installation status and a description.

- Click **Install** button. Skyline will 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 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 **Yes** to install the packages.
- Once the installation is complete, the MSstats tools will appear in the **External Tools** form.



Note: Your External Tools menu may already contain some Tools

- Click to **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**. 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.

Annotating samples with group information

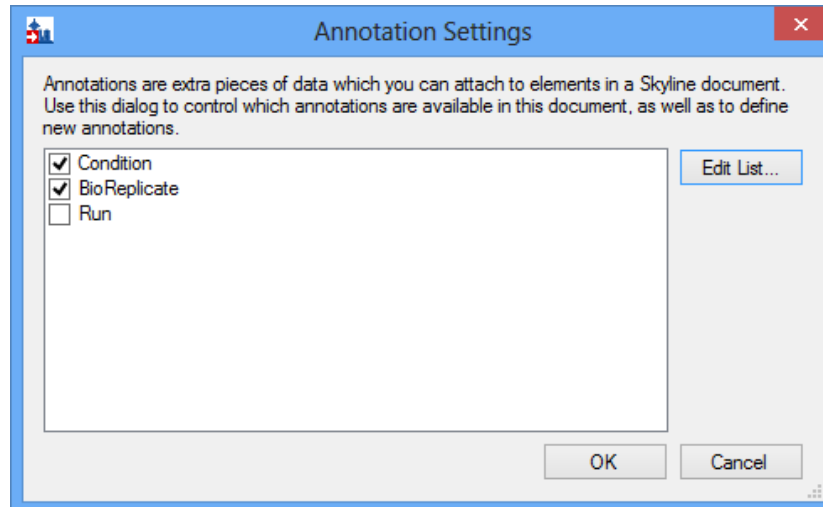
Skyline allows you to associate additional information with the replicates in the document by defining custom annotations. When `Human_plasma.zip` file was opened, it should have defined two custom annotations: **Condition** and **BioReplicate**.

The **Annotation Settings** form will show the list of annotations that have been defined.

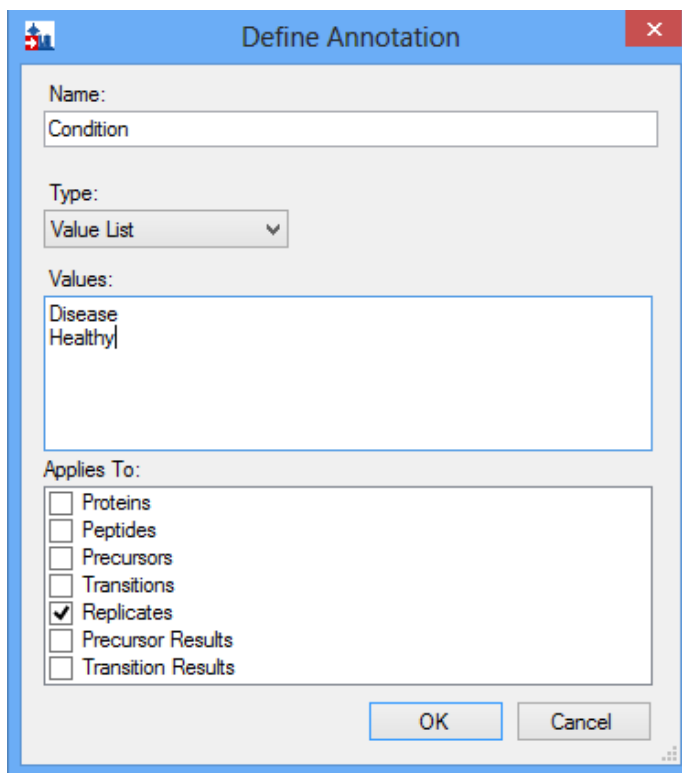
To view the **Annotation Settings** form, perform the following step:

- On the **Settings** menu, click **Annotations**.

You must check the checkboxes in the list in order to be able to use these annotations in your current Skyline document. The **Annotation Settings** form should look like the one shown below:



- Click the **Edit List..** button.
- Select **Condition** on Annotations. Then click the **Edit..** button.
- Check whether values of Condition in the **Values:** box are correct.



- Click the **OK** button of the **Define Annotation** form and **Annotation Settings** form.

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** (Alt-2).

The **Results Grid** will show you chromatogram peak areas and other measured results for the currently selected peptide, or transition. Opening the specified `Human_plasma.zip` file should have populated the results grid. Ensure that the data in the results grid matches the following:

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.

Results Grid			
Views	1	of 81	Export... Find:
Replicate	Condition		BioReplicate
A1	Disease	▼	1
A3	Disease	▼	2
A4	Disease	▼	3
A5	Healthy	▼	4
A6	Disease	▼	5
A7	Disease	▼	6
A8	Disease	▼	7
A9	Disease	▼	8
A10	Disease	▼	9
A11	Disease	▼	10
A12	Disease	▼	11
B1	Disease	▼	12
B2	Disease	▼	13
B3	Disease	▼	14
B4	Healthy	▼	15
B5	Disease	▼	16
B6	Disease	▼	17
B7	Disease	▼	18
B8	Disease	▼	19
B10	Disease	▼	20
B11	Disease	▼	21
B12	Healthy	▼	22
C1	Disease	▼	23
C2	Disease	▼	24
C3	Disease	▼	25
C4	Disease	▼	26
C5	Disease	▼	27

Results Grid			
Views	1	of 81	Export... Find:
Replicate	Condition		BioReplicate
C6	Healthy	▼	28
C7	Disease	▼	29
C8	Healthy	▼	30
C9	Healthy	▼	31
C10	Disease	▼	32
C11	Disease	▼	33
C12	Healthy	▼	34
D1	Disease	▼	35
D2	Disease	▼	36
D4	Disease	▼	37
D5	Disease	▼	38
D6	Healthy	▼	39
D7	Disease	▼	40
D8	Disease	▼	41
D9	Disease	▼	42
D10	Healthy	▼	43
D11	Disease	▼	44
D12	Disease	▼	45
E1	Disease	▼	46
E2	Disease	▼	47
E3	Disease	▼	48
E4	Healthy	▼	49
E5	Disease	▼	50
E6	Disease	▼	51
E7	Disease	▼	52
E8	Disease	▼	53
E9	Disease	▼	54

Results Grid

Views ▾ | 1 of 81 | Export... | Find: ▾

Replicate	Condition	BioReplicate
E10	Disease	55
E11	Disease	56
E12	Disease	57
F1	Disease	58
F2	Disease	59
F3	Disease	60
F4	Disease	61
F5	Healthy	62
F6	Healthy	63
F7	Healthy	64
F8	Disease	65
F9	Disease	66
F10	Disease	67
F11	Disease	68
F12	Disease	69
G1	Healthy	70
G2	Disease	71
G3	Healthy	72
G4	Disease	73
G5	Disease	74
G6	Disease	75
G7	Disease	76
G8	Disease	77
G9	Disease	78
G10	Disease	79
G11	Disease	80
G12	Disease	81

- Close the **Results Grid** in Skyline by clicking the red X in its upper right corner.

Using MSstats

MSstats is composed of three individual tools:

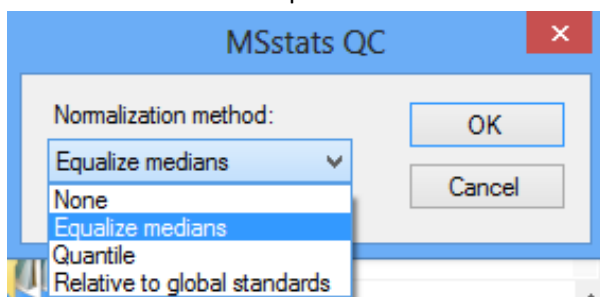
- **QC:** Provides quality control statistics for MS runs that contain the quantitative data for next analysis step. The step applies a logarithm base 2 transform, as well as a normalization that removes systematic differences in intensities between the MS runs.
- **Group Comparison:** Uses a family of linear mixed-effects models to test the protein abundances for significant changes across conditions.
- **Design Sample Size:** Viewing the current dataset as a pilot experiment, calculates the sample size for future experiments that would utilize the same experimental conditions.

For more detailed information about functionality and options, please visit msstats.org.

Running a QC Analysis:

To run a QC analysis, perform the following steps:

- On the **Tools** menu, choose **MSstats** and click **QC**.
- Choose normalization option on **Normalization method**. Here we will choose **Equalize medians**.



Skyline will begin exporting the MSstats Input report. Once the report is exported, Skyline will output the results in the **Immediate Window** docked at the bottom of the main Skyline Window. The QC step of MSstats will be launched automatically. Once this step is completed, navigate to the directory containing the file `Human_plasma.zip`. The directory should now contain a subdirectory called `Human_plasma`, with pdf files **ProfilePlot**, **QCPlot**, **ConditionPlot**, as well as the log file `msstats.log` file generated by the QC tool.

Running a Group Comparison:

To run a Group Comparison analysis, perform the following steps:

- On the **Tools** menu, choose **MSstats** and click **Group Comparison**.

Skyline will quickly display the **MSstats Group Comparison** form:

- Choose normalization option on **Normalization method**. Here we will choose **Equalize medians**.

- In the **Name of comparison** textbox, enter “Disease-Healthy”.
- Choose **Healthy** in **Control group**.
- Select other options as you would like.

- Click the **OK** button of the **MSstats Group Comparison** form.

The Group Comparison tool will now perform the analysis. The output from the tool run is shown in the **Immediate Window**, the textbox docked at the bottom of the main Skyline window. When the group comparison run is completed, switch back to the `Human_plasma` subdirectory again . The directory should now contain pdf files **VolcanoPlot**, **ComparisonPlot**, **ResidualPlot**, **QQPlot_allFeatures**, **QQPlot_byFeatures**, a csv file `TestingResult` with numeric values of the result, and `msstats-1.log` with the record of the analysis steps generated by the tool.

Running a Design Sample Size Analysis

To run a Group Comparison analysis, perform the following steps:

- On the **Tools** menu, choose **MSstats** and click **Design Sample Size**.

Skyline will quickly show the **MSstats Design Sample Size** form:

- Choose normalization option on **Normalization method**. Here we will choose **Equalize medians**.
- One of Sample size, Peptides per protein, Transitions per peptide, and Power needs to be selected to calculation. Other three values should be provided.

MSstats Design Sample Size

Normalization method: Equalize medians

Allow missing peaks

Automatically calculate

Sample size

Peptides per protein: 2

Transitions per peptide: 3

Power: 0.80

FDR: 0.05

Desired fold change

Lower: 1.25 Upper: 1.75

Use Defaults

OK Cancel

- For this tutorial you will leave the options in their default settings.
- 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**, the textbox docked at the bottom of the main Skyline window. When the tool is completed, switch back to the `Human_plasma` subdirectory again. The directory should now contain a pdf file **SampleSizePlot**, a csv file **SampleSizeCalculation** with numeric values of the result, and `msstats-2.log` with the record of the analysis steps generated by the tool.