

MSstats Group Study Statistics

This tutorial will guide you through the installation of MSstats and using it to perform statistical analysis on a sample Skyline file.

Getting Started

To start this tutorial, download the following ZIP file:

<http://skyline.gs.washington.edu/tutorials/MSstatsTutorial.zip>

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 files:

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

MSstats-1_0.zip: The MSstats external tool.

You will use the Human Plasma dataset for illustration. To open this dataset perform the following steps:

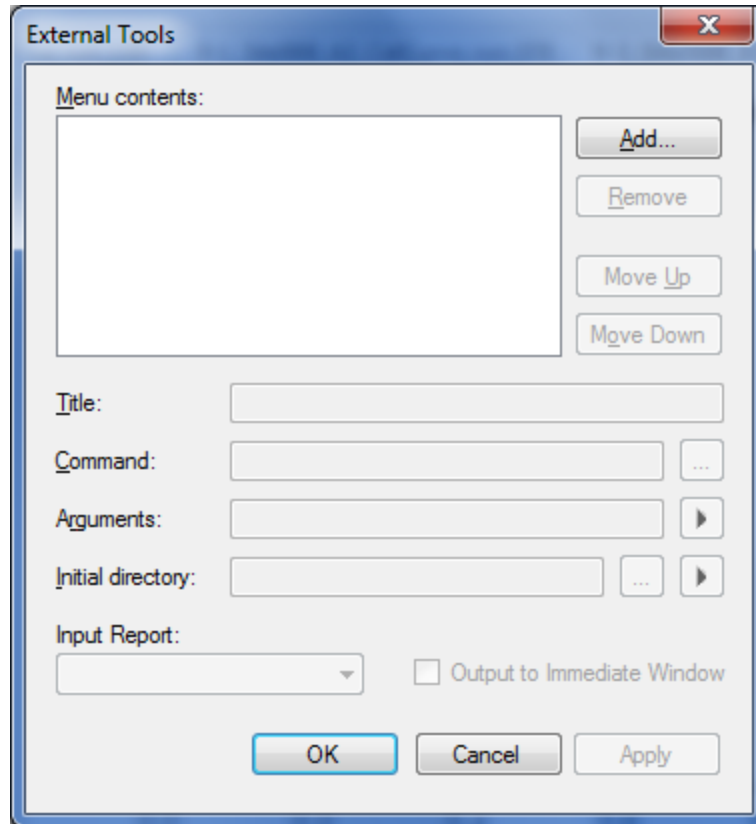
- Open **Skyline**.
- On the **File** menu, click **Open**.
- Change the **file type** in the **File name** field from 'Skyline Documents (*.sky)' to 'Shared Files (*.zip)'
- Navigate to the 'Human_plasma.zip' file in the folder you created and open it.

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:

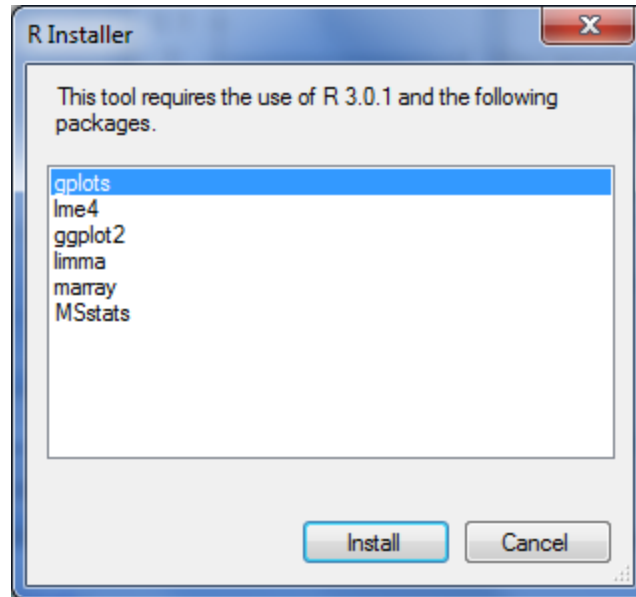
- On the **Tools** menu, click **External Tools** to bring up the **External Tools** form.



Note: Your External Tools menu may already contain some Tools

- Click the **Add** button, and select **From File...** from the menu that appears.
- Navigate to the location of the `MSstats-1_0.zip` file in the folder you created and select it.
- Click **Open**.

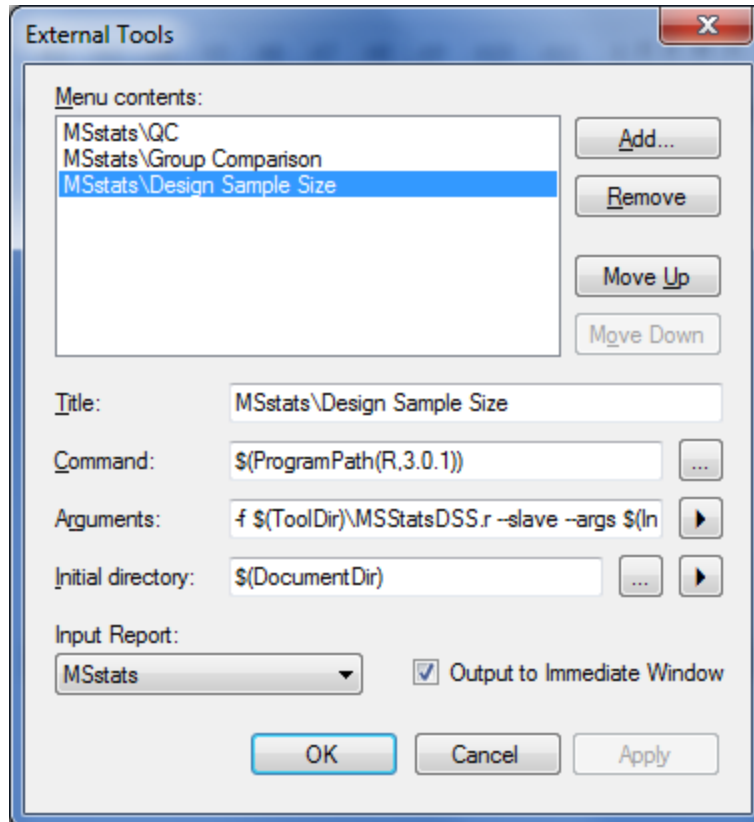
Skyline will now guide you through the process of installing MSstats as an external tool. This process may require you to install the R statistical programming environment, as well as a number of R packages which are used by MSstats to perform its statistical analysis. For example, if you do not have the appropriate version of R installed, the following form will be displayed:



To complete the installation, performing the following steps:

- On the **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 packages.

Once the installation is complete, the MSstats tools will appear in the **External Tools** form:



- Click to **OK** button on the **External Tools** form.

Annotating samples with group information

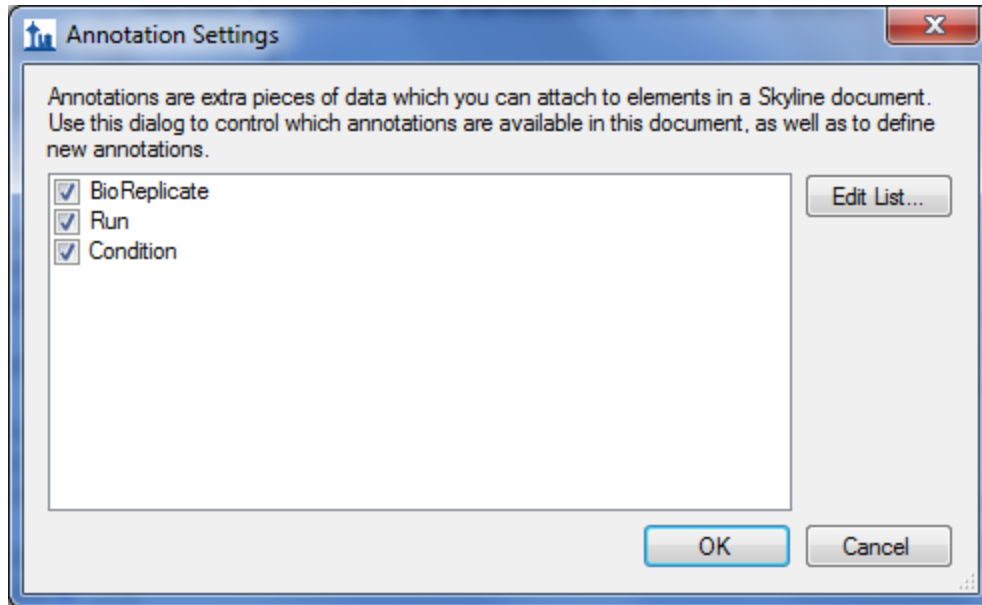
Skyline allows you to associate additional information with the replicates in the document by defining custom annotations. When the `Human_plasma.zip` file was opened, it should have defined three custom annotations: **Condition**, **BioReplicate** and **Run**.

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 **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** (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:

	Replicate Name	BioReplicate	Run	Condition
▶	A1	1	1	Disease
	A3	2	2	Disease
	A4	3	3	Disease
	A5	4	4	Healthy
	A6	5	5	Disease
	A7	6	6	Disease
	A8	7	7	Disease
	A9	8	8	Disease
	A10	9	9	Disease
	A11	10	10	Disease
	A12	11	11	Disease
	B1	12	12	Disease
	B2	13	13	Disease
	B3	14	14	Disease
	B4	15	15	Healthy
	B5	16	16	Disease
	B6	17	17	Disease
	B7	18	18	Disease
	B8	19	19	Disease
	B10	20	20	Disease
	B11	21	21	Disease
	B12	22	22	Healthy
	C1	23	23	Disease
	C2	24	24	Disease
	C3	25	25	Disease
	C4	26	26	Disease
	C5	27	27	Disease
	C6	28	28	Healthy
	C7	29	29	Disease
	C8	30	30	Healthy
	C9	31	31	Healthy
	C10	32	32	Disease
	C11	33	33	Disease
	C12	34	34	Healthy
	D1	35	35	Disease
	D2	36	36	Disease
	D4	37	37	Disease
	D5	38	38	Disease
	D6	39	39	Healthy
	D7	40	40	Disease
	D8	41	41	Disease

Record: < < 1 of 81 > > Filter:

Retention Times Results Grid Immediate Window

	Replicate Name	BioReplicate	Run	Condition
	D9	42	42	Disease
	D10	43	43	Healthy
	D11	44	44	Disease
	D12	45	45	Disease
	E1	46	46	Disease
	E2	47	47	Disease
	E3	48	48	Disease
	E4	49	49	Healthy
	E5	50	50	Disease
	E6	51	51	Disease
	E7	52	52	Disease
	E8	53	53	Disease
	E9	54	54	Disease
	E10	55	55	Disease
	E11	56	56	Disease
	E12	57	57	Disease
	F1	58	58	Disease
	F2	59	59	Disease
	F3	60	60	Disease
	F4	61	61	Disease
	F5	62	62	Healthy
	F6	63	63	Healthy
	F7	64	64	Healthy
	F8	65	65	Disease
	F9	66	66	Disease
	F10	67	67	Disease
	F11	68	68	Disease
	F12	69	69	Disease
	G1	70	70	Healthy
	G2	71	71	Disease
	G3	72	72	Healthy
	G4	73	73	Disease
	G5	74	74	Disease
	G6	75	75	Disease
	G7	76	76	Disease
	G8	77	77	Disease
	G9	78	78	Disease
	G10	79	79	Disease
	G11	80	80	Disease
	G12	81	81	Disease

Record: < < 1 of 81 > > Filter:

Retention Times Results Grid Immediate Window

- 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 become quantitative data for next step analysis. Logarithm transformation with base 2 and then constant normalization to remove systematic bias between MS runs is automatically 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.

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**.

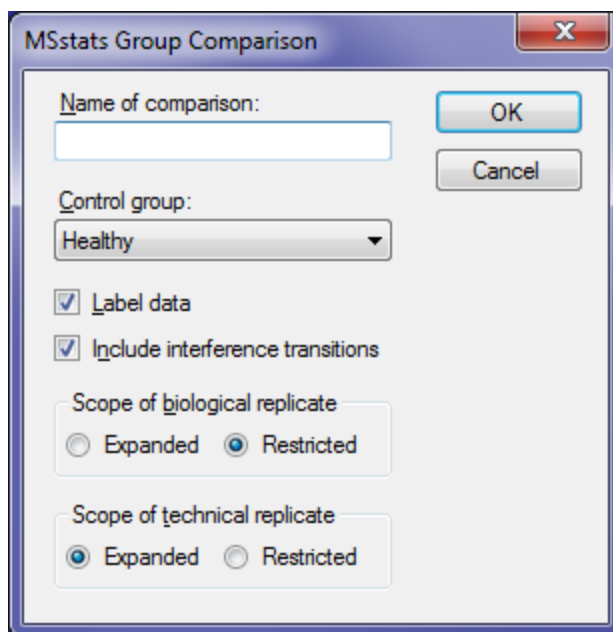
Skyline will begin exporting the MSstats Input report. This may take a few minutes. You can track the progress of the report export with the green progress bar in the bottom of the Skyline Window. Once the report is exported, Skyline will output the results of the tool run in the **Immediate Window** docked at the bottom of the main Skyline Window. When the tool is completed, navigate to the directory containing the `Human_plasma.zip` file. The directory should now contain a subdirectory called `Human_plasma`. This directory should now contain **ProfilePlot** and **QCPlot** PDF files 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**.

Because the Group Comparison tool uses the same report as the QC tool, and Skyline keeps recently used reports in memory, you will not see any report export progress this time. Skyline will quickly display the **MSstats Group Comparison** form:



- In the **Name of comparison** textbox enter 'Tutorial'.
- Click the **OK** button of the **MSstats Group Comparison** form.

The Group Comparison 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 group comparison run is completed, switch back to the `Human_plasma` subdirectory again. The directory should now contain **VolcanoPlot** and **ComparisonPlot** PDF files 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**.

The same report is used again for this analysis, and Skyline will quickly show the **MSstats Design Sample Size** form:

The image shows a dialog box titled "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. The first section, "Automatically calculate", contains three radio buttons: "Sample size" (which is selected), "Peptides per protein", and "Transitions per peptide". Below "Peptides per protein" and "Transitions per peptide" are spin boxes with the value "1". Below "Power" is a text box with the value "0.80". The second section, "FDR:", has a text box with the value "0.05". The third section, "Desired fold change", has two text boxes: "Lower:" with the value "1.25" and "Upper:" with the value "1.75". On the right side of the dialog, there are two buttons: "OK" and "Cancel". At the bottom center, there is a button labeled "Use Defaults".

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 **SampleSizePlot** PDF file generated by the tool.