

MS1Probe – Statistical Processing for MS1 full scan data

Developed by Alexandria K. Sahu, Birgit Schilling, Bradford W. Gibson (Buck Institute)

This tutorial will guide you through the **installation of MS1Probe** and subsequently guide you to perform **statistical analysis on a sample Skyline file**. The underlying statistical processing uses Python programming scripts and related Python modules which all will be automatically installed as part of the MS1Probe external tool installation.

MS1Probe is a tool that can further process MS1 full scan (MS1 Filtering) data originally analyzed in Skyline, and is designed to be capable of high throughput statistical quantification of Skyline MS1 Filtering datasets. Features of MS1Probe include calculating peak area means, variability measures, ratios between different sample conditions and corresponding q values and p values. MS1Probe arranges the MS1 data into a data array selected by replicate name and precursor ion m/z for M, M+1, M+2. MS1Probe then calculates the mean peak area, standard deviation, coefficient of variation (CV), peak area ratios, student's T-test p value, and q value for each peptide. The output contains all of this data in one comma separated variable report (csv). MS1Probe was constructed such that it can also process any MS2-based datasets (i.e., SRM, SWATH, MRM-HR).

In order to use MS1Probe with Skyline, it must be installed through the Skyline external tools framework. Furthermore, MS1Probe requires the use of custom annotations in Skyline. This tutorial covers the installation of MS1Probe, how to input data for custom annotations, and how to run the tool itself.

Getting started

To start this tutorial, download **MS1Probe_tutorial2.zip** from this page:

<https://skyline.gs.washington.edu/labkey/skyts/home/software/Skyline/tools/toolDetails.view?name=MS1Probe>

Extract the files in it to a folder on your computer, like:

C:\Users\bschilling\Documents

This will create a new folder:

C:\Users\bschilling\Documents\MS1Probe_Tutorial

The zip file contains the following files:

MS1_Probe_test_WT_KO_final.zip: Contains a Skyline document with sample data for MS1Probe.

You will use an MS1 Filtering test dataset containing *E. coli* wild type and mutant strain data 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 'MS1_Probe_test_WT_KO_final.zip' file in the folder you created and open it.
- Alternatively, you can also extract the zip file first and then open the *.sky file directly.

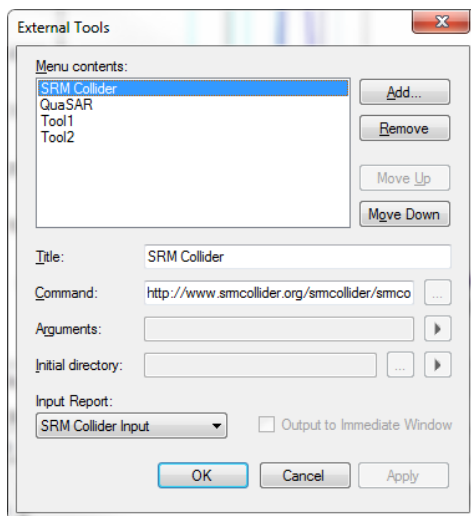
Installing MS1Probe as an External Tool

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

To add MS1Probe as an external tool, perform the following steps:

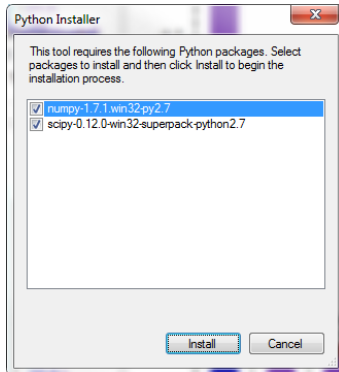
- On the **Tools** menu, click **External Tools** to bring up the **External Tools** form.

In the External Tools form you will either see no entries at all or a few preinstalled tools (i.e., SRM Collider) or your own tools that you may already have. Your External Tools form may look similar to this:



- Click the **Add** button, and select **From File...** from the menu that appears.
- Navigate to the location of the MS1Probe.zip file in the folder you created and select it.
- Click **Open**

Skyline will now guide you through the process of installing MS1Probe as an external tool. This process may require you to install the Python programming environment (currently Python 2.7, 32-bit will be installed even if a 64-bit computer is used due to Python compatibility aspects), as well as a number of Python extension packages, which are used by MS1Probe to perform its statistical analysis. You will be guided automatically through all of these steps. For example, if you do not have the appropriate 'NumPy' and 'SciPy' libraries, the following form will be displayed:

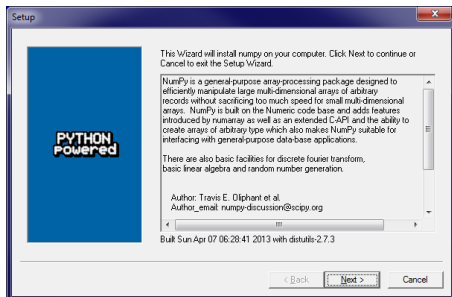


To complete the installation, performing the following steps:

- Keep numpy and scipy selected as suggested in the form, click **Install**.

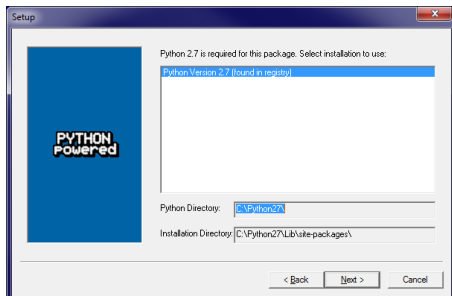
The download starts and you will be prompted with an Install Wizard for numpy.

Note: Look out for the Install Wizard window to appear (it might minimize automatically after initial appearance, in such case maximize again). You will see:



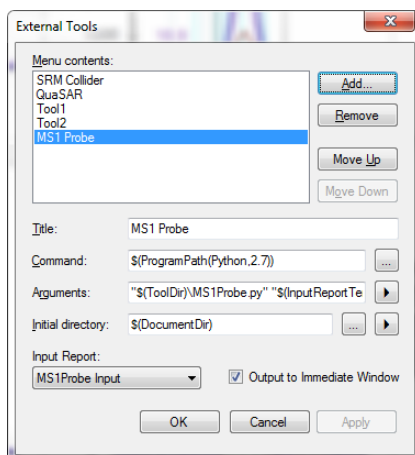
- Click **Next** to confirm

If Python 2.7 is already installed on your computer (and was found in the registry as shown below) you will be prompted with the following



- Then Click **Next** to confirm
- If you do not have the specified version of Python 2.7 installed on your machine, wait for Python to download. When the download is completed, you will be asked to install Python onto your computer. Complete the Python installation process by working through the Python installer wizard.
- Once the Python installation is completed or Python was initially already found in the computer registry, Skyline will continue to install the necessary Python packages, first numpy and then scipy, for MS1Probe. Continue to click **Next** and finally **Finish** when prompted by the Install Wizard. *Note: During the 'Installing Packages' steps Windows 7 may prompt you with a warning such as 'Do you want to allow the following program ... to make changes to this computer (it may only appear minimized on the task bar at the bottom of the screen, possibly flashing)', assess and allow to complete installation.*
- Finally, a window will appear saying 'Installation Complete' Click the **OK** button. Click OK even if you still see an 'Installing Packages' progress bar moving. The latter will disappear once OK is clicked.

Once the installation is complete, the MS1Probe tool will appear in the **External Tools** form:



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

Using the MS1Probe Tool to generate a statistical report

From your test Skyline file that you have currently open (MS1_Probe_test_WT_KO_final.sky) you will be guided to generate a statistical report from your MS1 full scan data.

Before starting, please read the **Overview** of MS1Probe functions first: Upon starting of the MS1Probe Tool within Skyline, several automatized steps will take place:

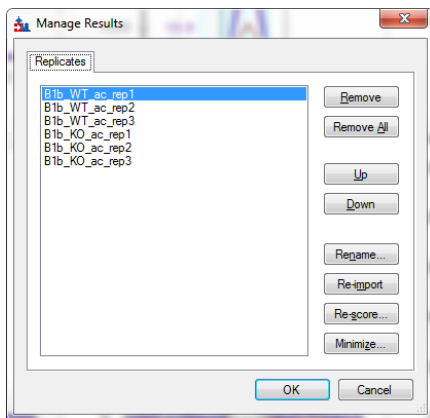
- Skyline will prompt the user in a command window to enter names of the two different conditions and/or sample types (i.e., cancer/control or wild type/mutant strain etc.) in the

datasets and the type of statistical peak area ratios (between the different conditions) that should be calculated.

- The MS1Probe tool extracts the data (MS1 precursor ion peak areas, XICs) from the automatically generated Skyline results report, and populates a data array for all MS1 precursor ions segregating replicates by custom conditions that the user had defined in the above steps (i.e. cancer/control type samples).
- Precursor Peak Area Means are calculated between the replicates of the specified condition, along with the standard deviation and the coefficient of variation. Ratios are calculated using the mean of the peak areas from all replicates per condition (per sample type, i.e., WT and KO, respectively) along with p-values to assess significance (Student's two-tailed t-test).
- Skyline/MS1Probe will prompt the user to append a name annotation to the generated Skyline report and the final Statistics Output. The two output files (*.csv) are generated and placed into the utilized Skyline File (*.sky) source folder. The Statistics Output file will automatically be opened in Excel.

As you saw distinguishing conditions, i.e. cancer/control, will have to be defined in the sample replicate name within Skyline (replicate names can be adjusted/changed in Skyline to generate distinct differentiators for the two different conditions to be statistically assessed by MS1Probe). You will need a minimally distinguishing description per 'condition' (which is case-sensitive), for example for the provided Skyline file familiarize yourself with the replicate nomenclature, there are three technical replicates of an *E. coli* wild type (WT) strain and three technical replicates of an *E. coli* knockout (KO) strain.

- In Skyline, on the **Edit** menu, click **Manage Results**.

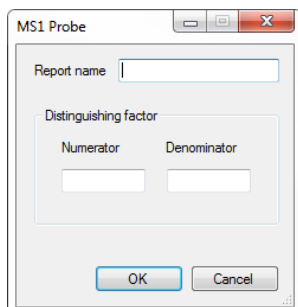


In this example the original MS file and replicate name already contained distinguishing vocabulary as part of the full Skyline replicate name that clearly differentiated the two conditions, here 'WT' for wild-type samples and 'KO' for knock out samples distinguish the two sample groups clearly. If your replicate naming happens to be too complicated and does not distinguish conditions you can change replicate naming with the **Rename** button in the above **Manage Results** form (if necessary) – changing replicate names in Skyline will not change original raw data file names (so no raw data naming gets manipulated).

In this case click on the **OK** button of the **Manage Results** form, and **start** the **MS1Probe tool** statistical processing:

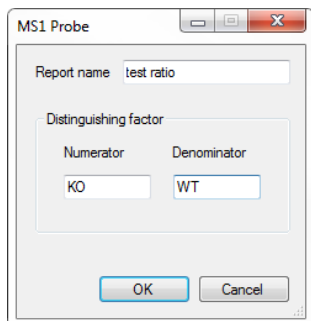
- In Skyline, from the **Tools** Menu, select the **MS1Probe** option.

The following **MS1Probe** form will appear:



- For Report name enter 'Test ratio'
- For the Distinguishing factor enter 'KO' as Numerator and 'WT' as Denominator (note entries are case sensitive).

Your MS1Probe form should look as follows



- Click the **OK** button

Your statistical report gets generated as a csv file, and opens immediately in excel. Expand several of the columns to view more descriptive details of headers or also of the transition column B to see the different extracted precursor ions M, M+1, and M+2 of the isotopic envelope.

The statistical report provides calculations for each of the 3 precursor isotopes, M, M+1, M+2 (rank 1, 2, 3, respectively) for WT peak area mean (including standard deviation and CV) and KO peak area mean (including standard deviation and CV). In addition, ratios as defined above by the user are calculated, i.e. ratio of peak area means for KO/WT with statistical significance measures such as a Student's t test p value and a further adjusted, false discovery based p value, referred to as q value (Storey, J. D. and R. Tibshirani (2003). "Statistical significance for genomewide studies." Proc Natl Acad Sci USA 100(16):

9440-9445). An example for an MS1Probe statistical output (after adjusting some peak area numbers for decimals in columns M through Q) is shown below for the first two peptides. Typically the highest ranked precursor ion provides the most useful output:

J	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	Peptide	Transition	Rank	LibraryRat	Modified Sequence	Precursor	Product M	Protein N	Protein D	Precursor	Product C	Retention	Mean KO	Stan Dev KO	CV KO	Mean WT	Stan Dev WT	CV WT	Ratio KO/WT	P-Value KO, WT	Q-Value K	Precursor Note	
2	MGANFLKVVNQI	precursor	1	#N/A	MGANFLK(+42)VVNGIK	752.4212	752.4212	sp	A72SL1	Elongation2	2	45.75	497776	37633	7.6	46182	5597	12.1	10.78	0.002916601	0.005066		
3	MGANFLKVVNQI	precursor [M+1]	2	#N/A	MGANFLK(+42)VVNGIK	752.4212	752.9227	sp	A72SL1	Elongation2	2	45.75	350842	30579	8.7	31076	3586	11.5	11.29	0.004116627	0.004561		
4	MGANFLKVVNQI	precursor [M+2]	3	#N/A	MGANFLK(+42)VVNGIK	752.4212	753.4235	sp	A72SL1	Elongation2	2	45.79	185267	14503	7.8	16079	3362	20.9	11.52	0.002456015	0.004561		
5	EEIKEVR	precursor	1	#N/A	EEIK(+42)EVR	472.756	472.756	sp	A72SL1	Elongation2	2	20.26	491494	8293	1.7	157889	3379	2.1	3.11	4.50E-05	0.000297		
6	EEIKEVR	precursor [M+1]	2	#N/A	EEIK(+42)EVR	472.756	473.2575	sp	A72SL1	Elongation2	2	20.26	243148	3060	1.3	86642	2667	3.1	2.81	8.39E-07	2.77E-05		
7	EEIKEVR	precursor [M+2]	3	#N/A	EEIK(+42)EVR	472.756	473.7587	sp	A72SL1	Elongation2	2	20.26	77614	2510	3.2	25936	3337	12.9	2.99	0.00010499	0.00041		

In excel filtering for the top ranked precursor (rank 1, column C), in this case precursor M, will simplify the obtained results; However, the availability of data for the first 3 precursor isotopes would allow you to address possible interferences by selecting the secondary ranked precursor ion.

Conclusion

In this tutorial, you have been introduced to install an external tool as well as to use a tool to statistically process MS1 Filtering (MS1 full scan) datasets using MS1Probe.

For technical questions or comments and suggestions please contact:

Birgit Schilling at bschilling@buckinstitute.org or

Alexandria Sahu at adsouza@buckinstitute.org

Using MS1Probe and Acknowledgments.

MS1Probe is provided as freely available tool within Skyline. All publications that utilize this software MS1Probe should provide acknowledgment to the Buck Institute for Research on Aging, the Gibson Laboratory/Chemistry Core and refer to the following original website link (<http://www.gibsonproteomics.org/resources/MS1Probe>).