Skyline Batch: A GUI tool for large-scale batch processing with Skyline

Alexandra N. Marsh[1], Vagisha Sharma[1], Surya Mani[1], Olga Vitek[1], Michael MacCoss[1], Brendan MacLean[1]

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, [2]Northeastern University, Khoury College of Computer Sciences

Introduction:

Skyline Batch is a newly developed Windows forms application that enables easy and consistent reprocessing of data with Skyline. By formalizing the workflow of a highly used set of batch scripts into an intuitive and powerful user interface, Skyline Batch can reprocess data stored in remote repositories just by opening and running a Skyline Batch configuration file. When run, a Skyline Batch configuration downloads all necessary remote files on its own and then runs a four-step Skyline workflow (**Figure 2**). By condensing the steps needed to reprocess the data into one file, Skyline Batch gives researchers the opportunity to publish their processing along with their data and other analysis files. These easily run configuration files will greatly increase the transparency and reproducibility of published work.

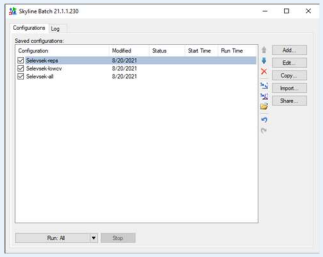


Figure 1. The main window of Skyline Batch. This is what first appears when the program is started. It contains a list of configurations on the first tab, and an output log on the secondary tab. Configurations represent instances of the workflow and can be run individually or in sequence.

Online Materials:

Download Skyline Batch

https://skyline.ms/batch.url

Skyline Batch Documentation: https://skyline.ms/batch-docs.url
Extensive documentation describing how to use every part of Skyline Batch to the fullest.
Skyline Batch webinar: https://skyline.ms/webinar20.url
A link to Webinar 20: Using Skyline Batch for Large-Scale DIA.
Skyline Batch written tutorials: https://skyline.ms/webinar14.url, https://skyline.ms/webinar15.url
Written tutorials for using Skyline Batch in webinars 14 and 15 are provided on the webinar pages.

Overview:

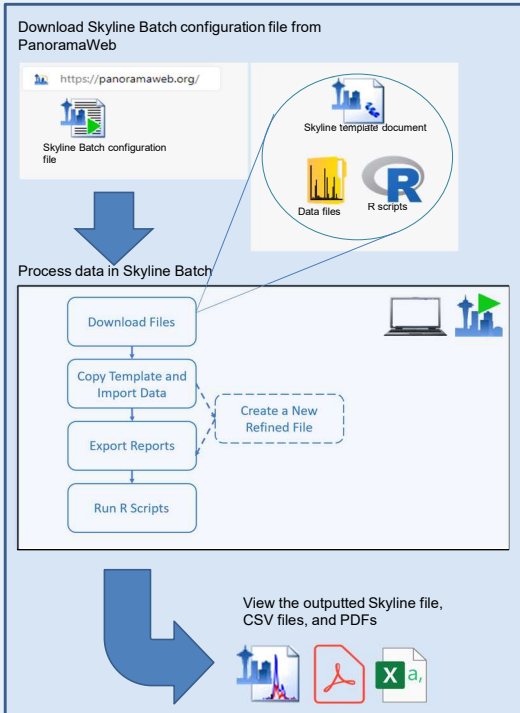


Figure 2. Reprocessing data from a remote repository in Skyline Batch. The Skyline Batch configuration file is downloaded and opened in Skyline Batch for processing. Skyline Batch downloads the Skyline template document, data files, and R scripts. The processing is run, and creates a Skyline results document, exported reports, and PDFs.

Sharing Batch workflows:

Skyline Batch was designed to make transferring batch workflows between computers as easy as possible. If a configuration (**Figure 3**) downloads the Skyline template document, data files, and R scripts from a Panorama server, all that the user must do to run the configuration on another computer is provide the path to a root folder in which downloads and processing will occur. In the scenario where some of the files were stored on disk on the original computer, Skyline Batch uses file path replacement to automatically guess where the same files will be stored on a different computer based on their location on the original computer.

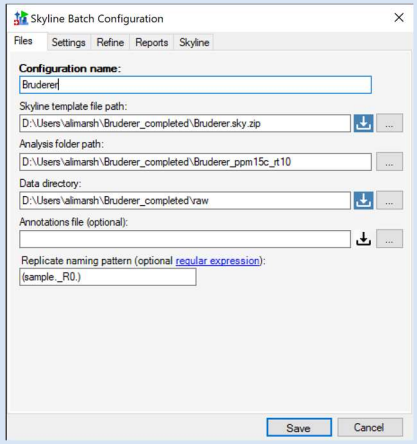
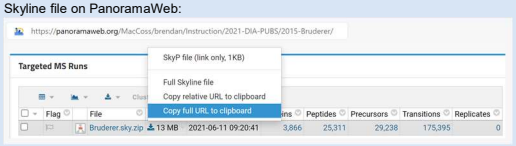


Figure 3. A Skyline Batch configuration on the "Files" tab of the edit configuration dialog. This configuration contains all the required information to run a workflow: the file locations of a Skyline template document, a folder containing data files, and any Skyline report files or R scripts that will be used for post-import results analysis.

Skyline file on PanoramaWeb:



Skyline file to be downloaded in Skyline Batch:

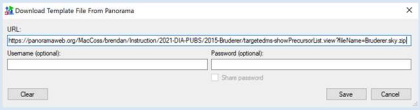


Figure 4. Selecting a Skyline template file to be downloaded in Skyline Batch. The first screenshot shows a page on PanoramaWeb containing a Skyline template file with the menu expanded. In the second screenshot from Skyline Batch, the link copied from the menu in the first screenshot is pasted into the URL field, indicating that this file will be downloaded when the corresponding configuration is run. For more information on downloading Skyline template files (and data files and R scripts) please visit the Skyline Batch documentation in the online materials section.

Results:

Skyline Batch fills the need for creating intuitive, automated Skyline workflows that can be easily executed by others. By connecting to Skyline and Panorama, Skyline Batch builds on top of tools that are already widely used in the proteomics community to make data processing easier and more transferrable. The GUI in Skyline Batch enables researchers, including those without scripting expertise, to configure reproducible Skyline workflows while capturing a record of the processing output in a log file. Skyline Batch configuration files can be shared to facilitate sharing of quantitative analyses of proteomics data and data reprocessing. Adding Skyline Batch to the Skyline software ecosystem fills a gap in scale and reproducibility for researchers lacking batch scripting and IT expertise.

Current work:

In current development, a Cromwell workflow will be designed to perform the processing specified by a Skyline Batch configuration file, enabling it to be run in a Docker Container on a server. This will further increase ease of access, as it will allow users to do the same batch analysis without downloading large data files onto their computers (**Figure 4**).