

# Fast-Track SWATH® Acquisition Data Processing with the SCIEX Cloud

## Overview of the OneOmics™ Project Proteomics Data Processing Pipeline

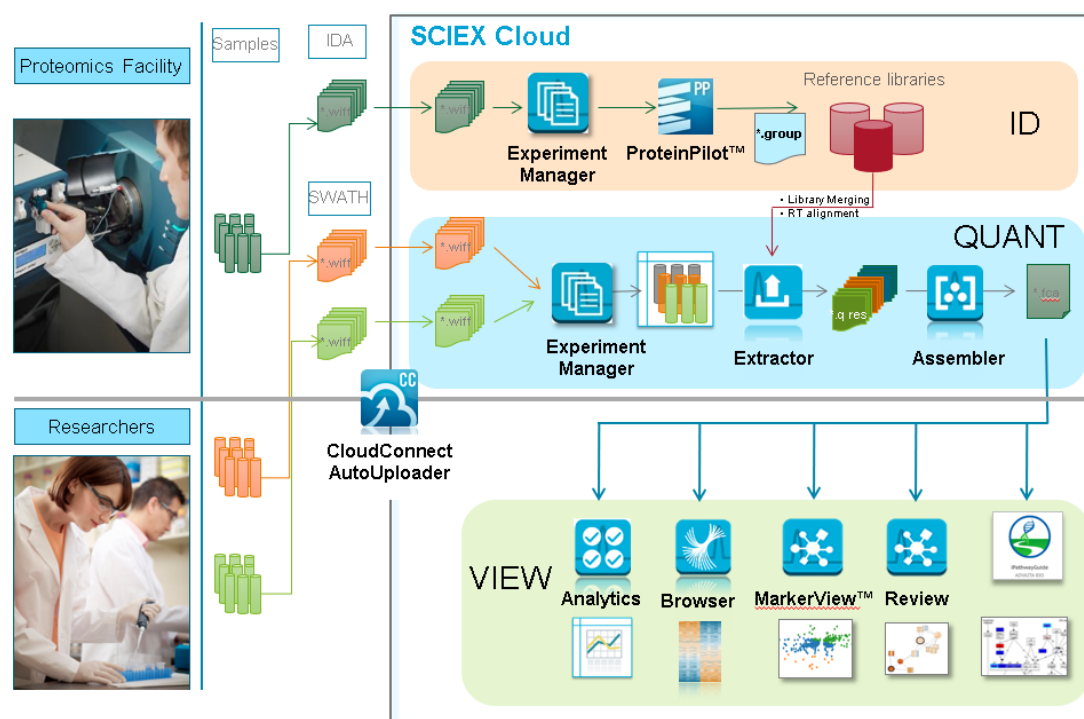
Because of the scale of the data that is now being created in Omics research, there is increasing pressure on data processing and an increasing need to move these intensive computations into cloud computing environments. This is the driver behind the OneOmics™ project at SCIEX. The SCIEX Cloud has been developed to provide storage and processing for multi-omics applications within a cloud-computing environment for easy, secure analysis and visualization of large and complex data sets.

Currently the SWATH Proteomics Cloud Toolkit consists of a series of applications to upload SWATH® Acquisition data to the cloud, process large datasets, then visualize and interpret the results (Figure 1). Results can also be securely shared with collaborators through specific sharing of results projects from one user account to another within the environment.



The processing and visualization applications allow researchers to view individual studies, as well as compare results across

multiple omics experiments and even align and compare results between RNA expression and proteomics experiments. Transcriptomics results can also be loaded (from either SCIEX data store or BaseSpace (Illumina)) into the visualization apps for comparison with proteomics expression data.



**Figure 1. Overview of the Processing and Visualization Pipeline.** Samples analyzed by SWATH® acquisition are uploaded into SCIEX cloud, then the metadata describing the study is specified in Experiment manager. The Extractor uses a spectral ion library to extract the peptide information from the data, followed by peak group scoring and FDR analysis. Ion Libraries can also be created in the SCIEX cloud by uploading DDA data and processing with ProteinPilot™ Software in the cloud. Next, data normalization and protein fold changes are computed in the Assembler application. Then data can be visualized with a variety of applications to assess MS data quality (Analytics), view protein expression differences (Browser), find trends in the data (MarkerView™ software), and assess the biological significance of the results (Review and iPathwayGuide (Advaita)).

## Overview of the SWATH® Acquisition Proteomics Toolkit

The following steps provide a high level overview of the steps to follow from data upload, through processing to visualization.

Click on the links in each section for more information on the various steps.

### A – Generate a Spectral Ion Library with ProteinPilot™ Software in the Cloud

1. Collect IDA files on pooled sample related to the samples in the study. Ideally some level of 2D fractionation is done on this pooled sample so a deeper ion library is created.
2. Ensure chromatography is stable such that retention times across the fractions are similar (no RT drift). Acquire data on all fractions in single batch to ensure RT reproducibility.
3. Data can be automatically uploaded to the cloud throughout the sample batch by using the [CloudConnect AutoUploader](#), to either a SCIEX Cloud or BaseSpace account.
4. Using [Experiment Manager](#), set up a study folder and then build a New Experiment. Select all the DDA data files to be processed in this experiment. Save the Experiment.
5. Return to the SCIEX Cloud dashboard by clicking on the SCIEX logo on the top left corner of the screen, then click on ProteinPilot.
6. Click the Process button, and then select Multi-File ProteinPilot Search to process all fractions together to create a single ion library. Select the Experiment that was defined above then click *Done*.
7. Provide a name for the processing step. Select a destination folder to save the processed results (\*.group).
8. Provide the processing settings and select a FASTA file.
9. Click *Execute* to start.

### B – Collect SWATH® Acquisition Data

Before starting the study, use the [SWATH Acquisition Performance kit](#) to confirm instrument performance.

1. Collect SWATH acquisition data files on all samples in the sample cohort.
  - Use 100 variable window method on most samples (25 msec accumulation time).
2. Run the PepCalMix 2-4 times per day to calibrate the MS throughout the study and monitor instrument performance.
3. Data can be automatically uploaded to the cloud throughout the sample batch by using the [CloudConnect AutoUploader](#), to either a SCIEX Cloud or BaseSpace account.

### C – Upload Ion Library to the Cloud

The Ion Library Group file must be first converted to a text file before being uploaded to the cloud. This must be done using [CloudConnect Software](#).

1. Import the ion library group file (filter by # of proteins at 1% global FDR) then load a representative SWATH file from the study.
2. Create a Retention Time Calibration Protein using endogenous peptides. (New [Auto Retention Time Calibration feature](#) in Extractor can also be used, making this step unnecessary).
3. Click Save, then select which data store to upload to. Specify the destination folder.

### D – Uploading SWATH Acquisition data to the Cloud

If AutoUpload wasn't used, the data files (\*.wiff) can be manually uploaded to the cloud using [CloudConnect Software](#). Data can be uploaded from a processing workstation or network drive.

1. Select Upload files to SCIEX or Upload files to BaseSpace to select the data files and the destination project in SCIEX Cloud or BaseSpace.
2. Monitor the status of upload in the upload queue.

### E – Define the MetaData using Experiment Manager

Now that the data and ion library are in the cloud, processing can begin. First step is to define the experimental metadata (experimental groups, biological and technical replicates) for the experiment.

1. From the SCIEX cloud dashboard, click on Experiment Manager.
2. Using [Experiment Manager](#), set up a study folder and then build a New Experiment. Select all the SWATH data files to be processed in this experiment.
3. Define the experimental groups and then assign the Experimental groups, biological and technical replicates for each data file. Every data file must have unique combination of experiment group, biological and technical replicates. Save the Experiment.

### F – Data Processing

The first step in data processing pipeline is to run the Extractor, which uses the ion library to extract out all the peak groups, then performs scoring and FDR analysis.

1. Return to the SCIEX Cloud dashboard by clicking on the SCIEX logo on the top left corner of the screen, then click on OneOmics.
2. Click the Process button, and then select [Extractor](#). Select the Experiment that was defined above then click *Done*.
3. Provide a name for the Extractor processing step. Select a destination folder to save the processed results (\*.qresults).
4. Select one or more ion libraries for processing ([Library merging feature](#)).
5. Provide the processing settings for Confidence, # of peptides and fragments, XIC width and RT window width.
6. Click *Execute* to start.

Next step is to run the [Assembler processing](#), where the data will be normalized using [MLR normalization](#) and the [protein fold changes](#) will be computed.

1. Select the Extractor results from the study.
2. Select a destination folder to save the processed results (\*.fcreports).
3. Define the attributes that will be the experimental groups, biological and technical replicates.
4. Click *Execute* to start.

## G – Data Visualization

Click on Workspaces to build the visualization dashboard for this set of processed results.

1. Click Create then select the Assembler results from the study, then provide a name for the workspace. If loading multiple results (multiple SWATH assembler results or SWATH and RNA results), experimental naming can be adjusted for consistency on the next tab.
- Use [Analytics](#) to review the MS data quality.
  - Use [Browser](#) to review the SWATH protein data in a biological context.

- Use [MarkerView software](#) to perform multivariate statistical analysis on the data sets.
- Comparative heat maps, enrichment analysis and relationship force-directed maps in Reviews provide power visualizations to relate and compare the data within the datasets.

## H – Using Data in Other Applications

The results output from Assembler are stored in the \*.fcreports file. This file can be downloaded and unzipped (change file extension to zip) to access all the underlying data and further interpret with other applications.

- The Foldchange[X].csv file can be directly uploaded into [iPathwayGuide \(Advaita\)](#) for pathway and ontology analysis.

## Important Links

1. SWATH Acquisition Performance Kit - <https://sciex.com/products/standards-and-reagents/swath-acquisition-performance-kit>
2. Get started by downloading the CloudConnect Software - <http://sciex.com/community/entity/9305>
3. View all the CloudTalks to learn more on each application - <http://sciex.com/community/entity/11173>
4. Read more details on the [algorithms and processing](#) in the OneOmics Project community. - <http://sciex.com/community/entity/1226>

AB Sciex is doing business as SCIEX.

© 2019 AB Sciex. For Research Use Only. Not for use in diagnostic procedures. The trademarks mentioned herein are the property of AB Sciex Pte. Ltd. or their respective owners. AB SCIEX™ is being used under license.

Document number: RUO-MKT-02-6969-B