

# Fast-track proteomics data processing with the SCIEX Cloud Platform

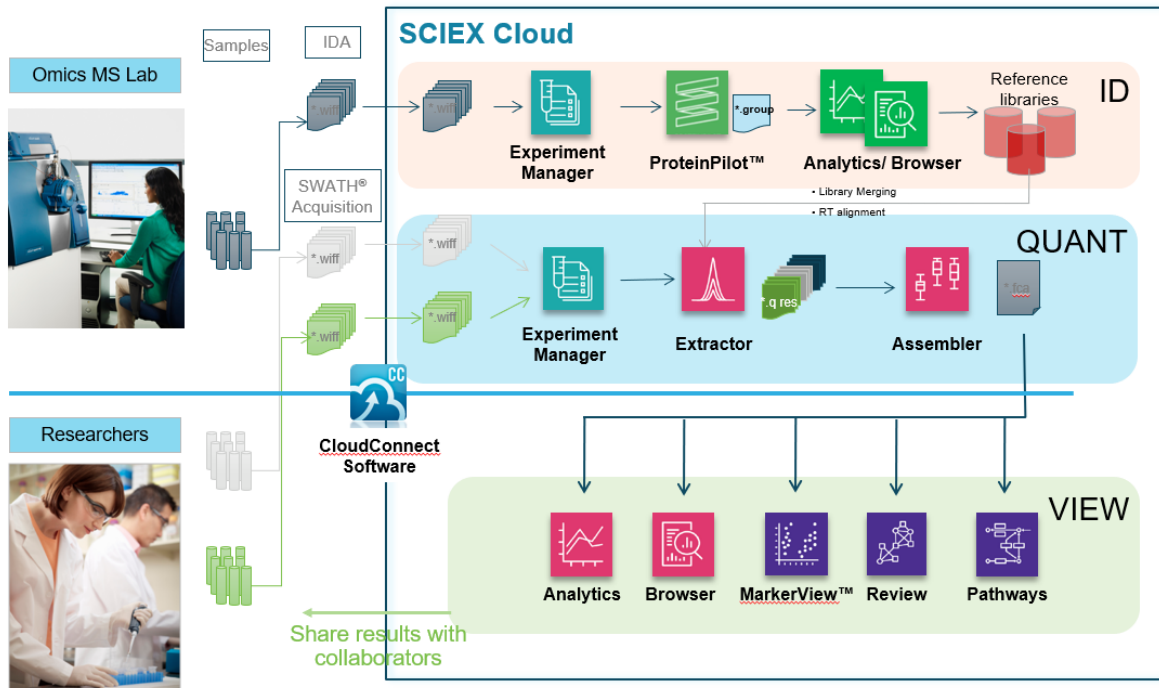
## Overview of the OneOmics™ Suite for omics data processing

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™ Suite at SCIEX. The SCIEX Cloud Platform 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 OneOmics Suite of applications consists of a series of applications to upload data to the cloud, process large datasets, and then visualize and interpret the results (Figure 1). Data can be uploaded to either the SCIEX Data Store or BaseSpace (Illumina). 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, proteomics and metabolomics experiments.



**Figure 1. Overview of the Processing and Visualization Pipeline.** Samples analyzed by the TripleTOF® System are uploaded into SCIEX Cloud Platform, then the metadata describing the study is specified in Experiment Manager. In the ID pipeline, protein identification results can be generated using ProteinPilot™ Software and visualized using Analytics and Browser. These results can also be used as ion libraries for subsequent SWATH® Acquisition data processing. In the quantification pipeline, first the Extractor app uses a spectral ion library to extract the peptide information from the data, followed by peak group scoring and FDR analysis. 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 Pathways). Results can then be easily shared back with collaborators.

## Overview of the workflow for OneOmics™ Suite

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. View the [Getting Started community post](#) to set up your account before you read on.<sup>2</sup> A demo dataset can be used to follow along with the full workflow.<sup>5</sup>

### 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 using the [CloudConnect Software](#) to either a [SCIEX Cloud](#) or [BaseSpace account](#).
4. Using [Experiment Manager](#), set up a study folder and then build a ProteinPilot ID Experiment. Select all the DDA data files to be processed in this experiment. Save the Experiment.
5. Return to the OneOmics Suite dashboard, then click on the [ProteinPilot Software](#) icon to begin data processing.
6. Click the Process button, and then select a Multi-File ProteinPilot Search to process all fractions together to create a single ion library. Select the Experiment that was defined above. Note a Per-file search can also be performed.
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.
10. Once the search is finished, you can view the results in Workspaces. Build a New Workspace in ProteinPilot and view your search statistics and protein ID results in [Analytics and Browser](#).

### 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. Upload data files to either a SCIEX Cloud or BaseSpace account.

### C – Upload ion library to the cloud

If using an ion library created on the desktop, it must first be 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) and 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. The first step is to define the experimental metadata (experimental groups, biological and technical replicates) for the experiment.

1. From the OneOmics Suite dashboard, click on Experiment Manager.
2. Using [Experiment Manager](#), set up a study folder and then build an Protein SWATH 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/or technical replicates](#) for each data file. Every data file must have unique combination of experiment group, biological and technical replicates. Save the Experiment.

## F – SWATH Acquisition 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 OneOmics Suite dashboard and then click on Proteomics
2. Click the Processing button, and then select [Extractor](#). Select the Protein SWATH Experiment that was defined above.
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. Click the Processing button, and then select Assembler.
2. Select the Extractor results from the study.
3. Select a destination folder to save the processed results (\*.fcreports).
4. Provide a name for the Assembler processing step. Define the attributes that will be the experimental groups, biological and technical replicates.
5. Click *Execute* to start.

## G –Data visualization

After every Assembler job, a workspace is automatically created. Click into Proteomics then click on Results on the left hand menu to view the processed results.

1. Click on the row of the results to view.
2. Use [Analytics](#) to review the MS data quality.
3. Use [Browser](#) to review the SWATH protein data in a biological context.

## H – Biological review

- Return to the main dashboard and click Biological Review or select from the left hand menu. Create a Workspace in Biological review using .fcreports output files from Assembler.
- Once Workspace is built, use the [MarkerView™ Software application](#) to perform multivariate statistical analysis on the data sets.
- Comparative heat maps, enrichment analysis and relationship force-directed maps in [Reviews](#) provide powerful visualizations to relate and compare the data within the datasets.
- Use Pathways to perform pathway analysis
- Note most tools within the Biological Review workspace can be used to perform [MultiOmics analysis](#).

## I – 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](#)
2. [Get started](#) here, learn how to set up your accounts, download the CloudConnect Software and access the demo data.
3. View all the [CloudTalks](#) to learn more on each application.
4. Read more details on the [algorithms and processing](#) in the OneOmics Suite community.
5. [Demo dataset](#) for trying OneOmics Suite in the SCIEX Cloud Platform.

The SCIEX clinical diagnostic portfolio is For In Vitro Diagnostic Use. Rx Only. Product(s) not available in all countries. For information on availability, please contact your local sales representative or refer to <https://sciex.com/diagnostics>. All other products are For Research Use Only. Not for use in Diagnostic Procedures.

Trademarks and/or registered trademarks mentioned herein, including associated logos, are the property of AB Sciex Pte. Ltd. or their respective owners in the United States and/or certain other countries. © 2020 DH Tech. Dev. Pte. Ltd. RUO-MKT-02-6969-B. AB SCIEX™ is being used under license.