Biomarkers and Omics

SCIEX
The Power of Precision

Fast-track proteomics data processing with the OneOmics suite

Omics data processing in the cloud

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. This 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
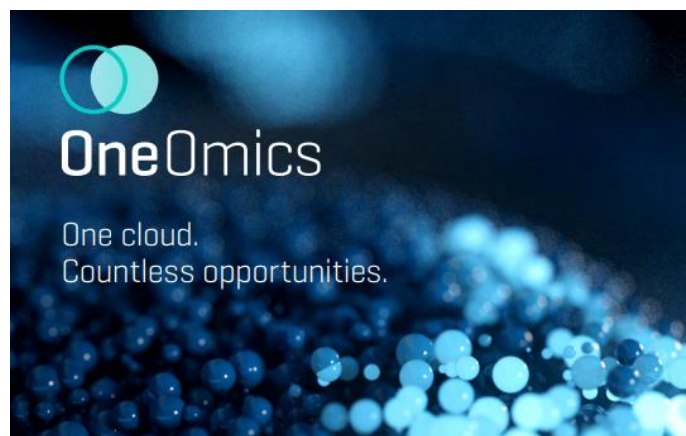
multiple omics experiments and even align and compare results between RNA expression, proteomics and metabolomics experiments.
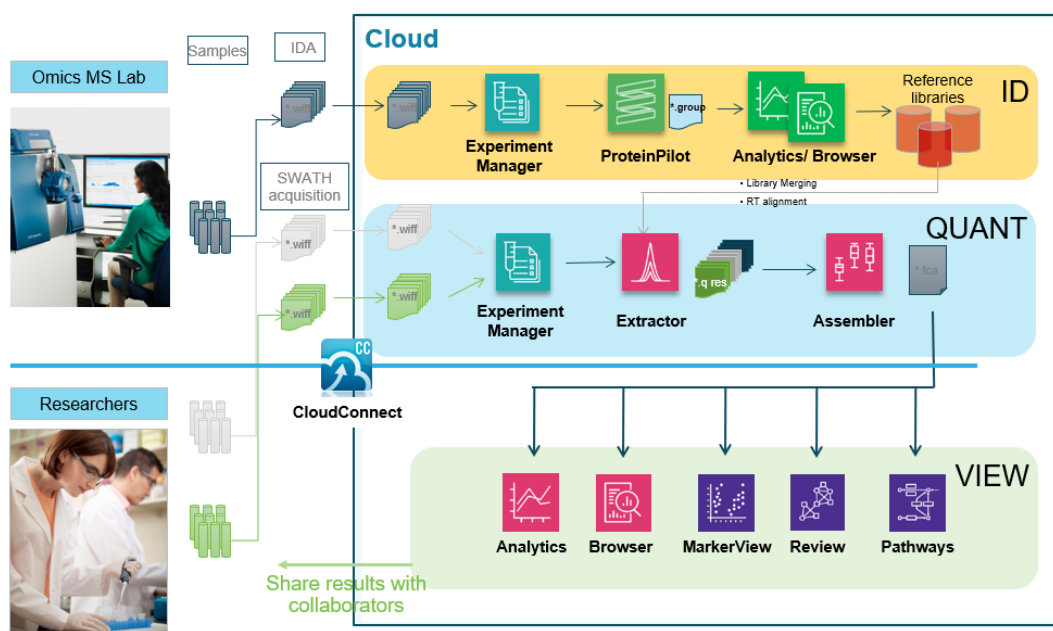

Figure 1. Overview of the processing and visualization pipeline. Data files are uploaded into the cloud, and then the metadata describing the study is specified in Experiment Manager. In the ID pipeline, protein identification results can be generated using the ProteinPilot App and visualized using Analytics 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.[1] A demo dataset can be used to follow along with the full workflow.[2]

A – Generate a spectral ion library with ProteinPilot app 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 uploaded to the cloud using CloudConnect to either a SCIEX Data Store 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 app 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.[3]

1. Collect SWATH acquisition data files on all samples in the sample cohort using an optimized method.

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 the SCIEX Data Store or a 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.

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

Data files (*.wiff) can be manually uploaded to the cloud using CloudConnect. 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 Data Store 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 a Protein SWATH Experiment. Select all the SWATH acquisition 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 a unique combination of experiment group, biological and/or 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 (*.fcresults).
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 acquisition protein data in a biological context.

H – BioReviews

- Return to the main dashboard and click BioReviews or select from the left hand menu. Create a Workspace in Biological review using .fcresults output files from Assembler.
- Once Workspace is built, use the MarkerView app 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 *.fcresults 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.

- For example, the Foldchange[X].csv file can be directly uploaded into iPathwayGuide (Advaita) for pathway and ontology analysis.

Important links

1. Get started here, learn how to set up your accounts, download the CloudConnect Software and access the demo data.

2. Demo dataset for trying OneOmics suite in the cloud.

3. SWATH acquisition Performance Kit

4. View all the CloudTalks to learn more on each application.

5. Read more details on the algorithms and processing in the OneOmics suite community.

Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
sciex.com

International Sales
For our office locations please call the division headquarters or refer to our website at
sciex.com/offices