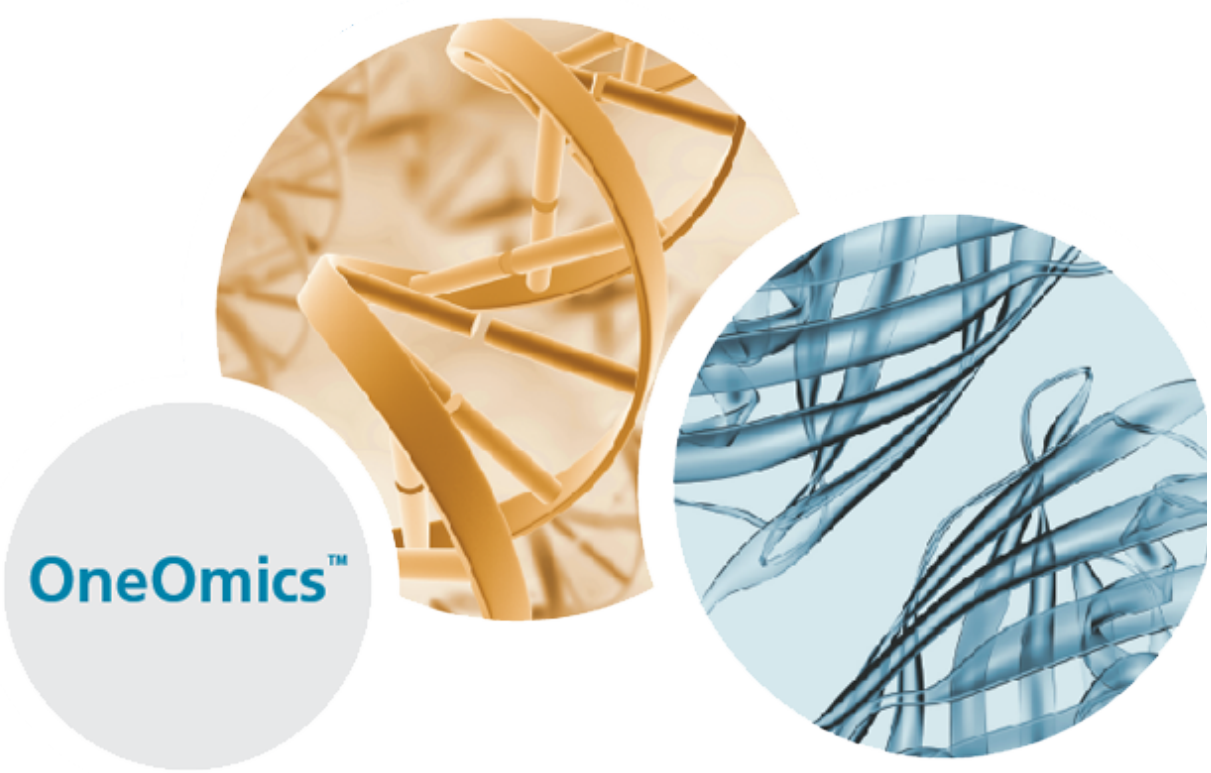
OneOmics™ Project 2.1 Beta - for SWATH® Acquisition Data Processing

Release Notes



RUO-IDV-03-2142-E

October 2017

AB Sciex Pte. Ltd.
Blk 33, #04-06
Marsiling Ind Estate Road 3
Woodlands Central Indus. Estate.
SINGAPORE 739256

Contents

New Features and Changes 1

This version of the OneOmics™ Project cloud delivers all of the applications that allow for the Web-based processing and analysis of data files acquired on any SCIEX TripleTOF® system (5600, 5600+, and 6600 systems) enabled with SWATH® acquisition. This release of the OneOmics Project is in beta phase, and therefore has not been fully validated. There are a few known issues as described in these release notes. Please use the Feedback button within the Cloud applications to provide suggestions and report any issues.

New Features in this version:

- The OneOmics™ Project cloud applications for the processing of SWATH® acquisition data have now been moved into the SCIEX Cloud.
- In addition, a new data storage capability has been developed that allows for storage of SWATH acquisition data within the SCIEX Cloud. (Please see section on known issues.)
- BaseSpace (Illumina) can still be used as a data store (for both SWATH Acquisition data storage and RNA-Seq data storage and processing) and is integrated seamlessly with SCIEX cloud for multi-omics processing. Please refer to the BaseSpace rules and security policies for storage limits and use of applications for processing within BaseSpace.

A detailed description of the workflow is described in the Help menu of each application. In summary, using the CloudConnect Micro-App, the user can upload a SWATH® acquisition data file and a text ion library to BaseSpace or SCIEX Data Store. Refer to the *CloudConnect Micro-App 1.2 Beta User Guide*. As an alternative to using the CloudConnect Micro-App, users can use the CloudConnect AutoUploader to automatically upload data, post-acquisition, from the instrument workstation. The CloudConnect AutoUploader monitors an Analyst batch and upon acquisition completion uploads the data file to the cloud. Refer to the *CloudConnect AutoUploader 1.2 Beta User Guide*.

After the data is uploaded the user sets up the experimental methodology in the Experiment Manager application. The next steps occur in the new OneOmics application for proteomics and transcriptomics processing. The users can process the SWATH® acquisition data using the Extractor application followed by the Assembler application to perform real-time fold change analysis. Results can be viewed in the Browser application for biological interpretation or the Analytics application for viewing MS data quality. The output of the Assembler fold change analysis is an .fcresult file and a .csv file containing fold change data. The .fcresult file is a compressed file that can be downloaded, renamed to .zip, unzipped and then used for downstream data visualization application tools. The .csv files can also be downloaded for use by other tools, or used by other BaseSpace applications, such as Advaita iPathwayGuide.

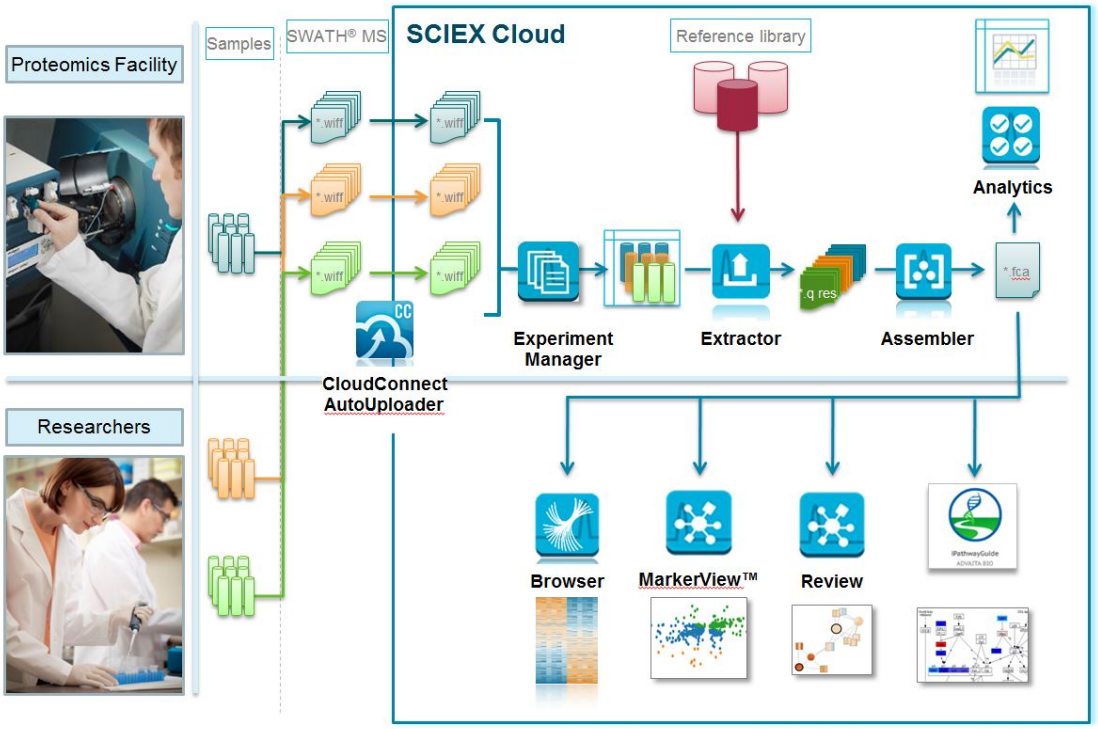
Enhancements delivered with this release include:

- The ability to Integrate multiple Omics studies, to find similar trends and differences.
    - Combine studies together to align the data
        - RNA-Seq vs proteomic
        - Proteomic vs proteomic
- Perform supervised and unsupervised multivariant analysis techniques to find similarities / differences between experiments and within experiments.
- Determine biological significance of these similarities / differences
- Datafiles, ion libraries, and processed results can be saved to SCIEX Cloud in addition to BaseSpace.

OneOmics™ Project Workflow

The OneOmics™ Project delivers a series of cloud applications (Figure 1-1) that enable the fast processing of SWATH® acquisition data in the cloud and the review of that processed data with a number of visualization applications.

Figure 1-1 OneOmics™ Project – Workflow Overview



1. Individual SWATH® acquisition mass spectrometry (.wiff) files and the ion library (.txt) file must be uploaded from the instrument computer or the processing computer to the cloud using either the CloudConnect Micro-App, available as an option for the PeakView® software, or the CloudConnect AutoUploader utility, available in the Analyst® TF software.

2. First, the Experiment Manager application is used to define the experimental metadata for experimental groups in a study.

3. The next steps occur within the OneOmics application.

    a. The Extractor application extracts and integrates the peptide or protein peaks using the ion library. A .qresult file is created as an output for every sample.

c. Next, the Assembler application uses the study design information from step #2 to normalize the data and compute the protein fold changes across the samples.

d. These protein fold changes can be visualized using the Browser application, across the study, across the protein sequence, within a biological context.

e. The quality of the underlying MS data can be viewed in the Analytics application. There are plots in the Analytics application that show the peak area %CV, the normalization quality, FDR plots, and many other useful visuals.

f. Next MarkerView(TM) can be used for various multivariate and statistical analysis techniques to find similarities and differences within an experiment or between your experiments.

g. And finally, powerful visulization tools in Reviews help you to relate and compare the data within your experiment sets.

New in the Experiment Manager Application

- More flexibility in metadata entry and experimental group naming.
- You can now add and delete samples when building experiments in the metadata editor.
- Data Store is the SCIEX cloud data storage, allowing you to upload and store your data within the SCIEX Cloud.
- You can mark an experiment as a Favorite which adds it to the Favorites list for easy access

New OneOmics Application

- Extractor, Assembler, Browser, Analytics, MarkerView™ and Reviews have all been combined into the OneOmics Application.

New in the Extractor Application

- Merging of multiple libraries is now supported.
- ProteinPilot 5.0 Group files can now be used for your ion library. Shared peptides are automatically excluded and modified peptides are included automatically.
- You can now do Auto-calibration which automatically selects the best compounds throughout the RT range in your library. (See the help menu for more details on auto-calibration).
- The **XIC extraction width** has moved into the Calibrate step of Extractor because this also sets the width of the calibration step.

New Assembler Application

- A new and improved Assembler application has been added.
  - Improvements have been made to the algorithm.

New Workspace Application

- The new Workspace application enables you to visualize the results within experiments and build multi-omic investigations to align and compare your data allowing for exploration of similarities and differences within an experiment or between the experiments.
- Workspace houses Analytics, Browser, MarkerView, Conflict Resolution and Reviews.

New Conflict Resolution Feature

- Conflict Resolution allows you to resolve a conflict when a protein in the protein results maps to multiple genes in the RNA results or when a single gene in the RNA results maps to multiple proteins in the protein results.

New in the Browser Application

- Multivariate analysis has been removed from the Browswer application and exists in the MarkerView Application only.
- For performance reasons we are now importing only one fcresult file into Browser therefore the ability to change the control in the heat map is no longer available.

New MarkerView Application

- A new and improved Markerview application has been added with multiple multi-variant analysis techniques (both supervised and un-supervised).
- Area-based and Fold-change based cluster analysis session are supported.

New Review Section

- The Reviews section is useful for sifting through your study's various data sets. Powerful visualizations help you to to more deeply explore your datasets, for comparing multiple datasets and exploring the ontologies implicated from the protein or gene expression data. Through these visualizations you can drill into areas you identify as interesting and explore other related data that may be relevant.

Known Issues 2

Note: The numbers in brackets are reference numbers to each issue or feature in the SCIEX internal tracking system.

- The Help menu content provides details on the workflow and mentions Metabolomics workflow. The Metabolomics workflow is not yet available in this Beta.
- Unlike all the other svg exports, the force-directed map svg export can't be opened in Chrome (it just shows as XML).
- Sciex Cloud only supports file names of 255 characters or less. Names that will not be supported are those that contain non-printable ascii, / or \, names with leading or trailing spaces, and the special names “.” and “..”.
- Uploading datafiles and saving processed results to SCIEX Cloud Data Store can be slower and less stable than uploading and saving to BaseSpace. Use BaseSpace to upload to and save to when processing large jobs (>30 samples).
- Non-UniProt identifiers are not supported. If the proteins in the ion library do not include UniProt identifiers, then the analysis might fail. (OOM-39)
- It may seem as if the application is not responding as there are delays in screen updates. To fix this click F5 to refresh the screen.
- Occasionally, the Extractor application analysis does not complete. To fix this issue, start a new analysis.
- The PCA-PCVG page from a single protein experiment session doesn't load in a multi-protein experiment workspace. If you want to perform PCA-PCVG on a single protein experiment then create a workspace with a single protein experiment.
- The search field in the Compare section is case sensitive. Use the necessary case to find what you are looking for.
- You cannot currently view a cluster session result in Compare.
- Viewing Analytics for two different results generated by extracting data from the same experiment with two different ion libraries will not work when you import both results into the same workspace session. You will only be able to view the results for one dataset. (For example, creating a workspace session with both CDK4 SWATH data extracted with the CDK4 library and CDK4 SWATH data extracted with PHL).
- RNAseq Cuffdiff files cannot be used as such. Rename the gene_exp.diff file as .txt and then import for use.

- Because we use Confidence instead of p-values to create a volcano plot, when viewing t-test results for a multi-omics experiment that includes RNA fold change data, the RNA data points will be on the x-axis.
- Reprocessing a reprocessed session is not working as expected in MarkerView. If you exclude a feature and then reprocess then exclude another feature (from the reprocessed results) and reprocess a second time the first feature will still be included. To work around this you will need to include all features to be excluded in each instance of reprocessing.
- Users should avoid using non-alphanumeric characters when naming an Experiment Group. If the Experiment Group name contains any special characters, such as a colon (:) or a backslash (\), then the generation of fold change analysis results for the experiment in the Protein Expression Assembler application might fail. (OOM-160)
- Results (.qresult files) that have been generated using a library that does not contain calibrated retention time information are missing the RT Metrics information in the Analytics application. The RT Metrics panel shows a deviation of 0s between peptide retention times and expected retention times. In addition, the retention time visualization page contains an error icon and the message "There were no expected retention times included with your library." instead of a plot showing the Observed RT versus Expected RT for each sample in the file. (OOM-187)
- The Principal Component Analysis (PCA) algorithm requires that the number of proteins in the experiment must be equal to or greater than the number of samples. In experiments where the number of proteins is less than the number of samples, the PCA cannot be performed and the Multivariant Analysis section of the Browser application is blank. (OOM-253)
- Depending on the number of proteins and contrasts included in a data set, the heat map page in the Proteins view of the Browser application may stop responding. We recommend that a maximum of 3,000 proteins across 15 contrasts be included, using the Protein Filtering Controls, to make sure that the generation and presentation of the corresponding heat map occurs in a timely manner. (OOM-255)
- The implementation of fixed column headers in the protein and peptide heat maps has resulted in the removal of the svg export icons (orange buttons in top right corner of plot windows). As a workaround, users can export the heat maps as .csv files and then open the file in Microsoft Excel and apply conditional formatting to achieve the required results. (OOM-316)
- Occasionally, in the Protein Sequence view, if the user hovers over a block of amino acids. a empty tool tip field might be shown. This occurs only when there is no additional relevant information that can be shown to the user.

Limitations

- The total number of transitions that the Protein Expression Assembler application can handle in one processing job is approximately 12 million after FDR filtering. In practical terms, this can be reached by exceeding the following scenarios:
  - Assembling 100 extracted result files (.qresult files) that were processed using a ProteinPilot™ software ion library containing 6,000 proteins.
  - Assembling 30 extracted results files .(qresult files) that were processed using an ion library with approximately 1 million transitions (for example, derived from the pan-human library).
- Multivariant grouping works well when most of the variance across samples can be explained by a small number of the principal components (PCs) identified in the Principal Component Analysis (PCA) algorithm. When there are many experimental groups with very high variance across samples, it is possible that it might take a large number of PCs to explain most of the variance. A negative consequence of this would be that a large number of groups can be formed, too many to be reviewed in the software. To fix this issue, we have limited the number of PCs that are used for grouping, even if this results in some PCs that explain a non-trivial amount of the variance being omitted. As a result, the number of groups formed will typically be in the interval [10,100], but minor differences in protein expression trends that ideally might be split into multiple groups are instead merged. If, after manual review, there are groups found where their member proteins exhibit dissimilar expression trends within the same group, then it might be possible to further split the groups by selecting the groups of interest and selecting **Reprocess**.

Revision History

Revision	Reason for Change	Date
A	First release of document.	March 2015
B	Added the New Features and Changes section, the Protein Expression Workflow section, detailing the features, enhancements, and application changes for this release. Moved issue OOM-51 to the New Features and Changes topic as a fixed issue. Added the following items to the Known Issues section: OOM-160, OOM-187, OOM-253, OOM-255, OOM-260, OOM-288, OOM-291, and OOM-316.	March 2016
C	Moved issue OOM-51 to end of New Features and Changes section.	May 2016
D	Add new features for 2.1 Beta. Updated workflow de-scription. Added new known issues.	October 2017
E	Added feature: ProteinPilot .group files can now be used as an ion library in Extractor. Added new known issues.	November 2017