

Integrating DIA-NN software analysis of SWATH data-independent acquisition (DIA) data into a cloud processing pipeline



Melanie Juba¹, Nick Morrice², Alexandra Antonoplis¹ and Christie Hunter¹
¹SCIEX, USA; ²SCIEX, UK

INTRODUCTION

The combination of omics disciplines has proven to be more powerful than individual disciplines. However, multi-omics data analysis workflows can be time-consuming, as consolidating and interpreting various results outputs – especially when using different scoring schemas and criteria – is often challenging. The cloud-based OneOmics suite was previously demonstrated to provide researchers tools for multi-omics data management, compound identification and quantitation, statistical analysis, and pathway analysis to streamline biomarker discovery studies¹⁻³.

To further facilitate data processing for high-throughput proteomics workflows, the OneOmics suite extends its SWATH DIA data processing by supporting visualization and statistical interpretation of DIA-NN software results. DIA-NN software is a widely used proteomics processing platform that leverages neural networks and powerful quantitation and inference algorithms to achieve confident protein and peptide identifications⁴.

Here, a processing workflow for proteomics analysis is presented using Zeno data-dependent acquisition (DDA) and SWATH DIA data. In this workflow, a spectral library was generated from Zeno DDA data using the Ion Library app in OneOmics suite. SWATH DIA data were subsequently processed using this ion library in DIA-NN software and evaluated in the DIA Results app (Figure 1). Results generated using this workflow can be rapidly and securely shared with collaborators⁵.

MATERIALS AND METHODS

Sample preparation

As previously described, a 100 µg sample of K562 cell lysate was fractionated using high-pH RP-HPLC⁶. For Zeno SWATH DIA on the ZenoTOF 7600 system, a sample of K562 digest (SWATH acquisition performance kit) was prepared in water with 1% formic acid and analyzed at loading amounts ranging from 12.5 – 200 ng. A data set consisting of SWATH DIA results corresponding to 6 human cell lines was also evaluated³.

Chromatography

Microflow LC analysis of the K562 fractions and 6 cell lysates was performed as previously described.^{3,6} For Zeno SWATH DIA experiments, a 45-minute microflow gradient from 5% to 30% mobile phase B on the Waters ACQUITY UPLC M-Class system was implemented in trap-elute mode with a Phenomenex C18 micro trap (10 x 0.3 mm) and a flow rate of 5 µL/min.

Mass spectrometry

A ZenoTOF 7600 system equipped with the OptiFlow Turbo V ion source using a low microflow probe and electrode was used for Zeno DDA and Zeno SWATH DIA data acquisition. For generation of the ion library, DDA parameters were implemented as previously described⁶. For Zeno SWATH DIA, an 80 variable window method was used with an MS/MS accumulation time of 25 ms and dynamic collision energy. The 6 cell lysates were analyzed as previously described on the TripleTOF 6600+ system³.

Data processing

DDA data files were uploaded to the OneOmics suite using CloudConnect in PeakView software, version 2.2. Data were then searched using the multi-file option in the Ion Library app using a human FASTA file from UniProt. This app creates ion libraries that are compatible with DIA-NN software. Search results were visualized using the Analytics and Browser apps in the ProteinPilot app to assess the quality of the protein identification results.

The generated ion library was then used to process SWATH DIA data in DIA-NN software, version 1.8.1. In DIA-NN software, the robust LC, high precision workflow setting was selected along with match between runs (MBR). The advanced command –report-lib-info was also added, as the –report-lib-info command is required to create an output file that is compatible with the DIA Results app in the OneOmics suite⁷. The results output from DIA-NN software consists of several *.tsv data files. The overall final *.tsv report for each search (the largest *.tsv file in size) was uploaded to OneOmics suite using CloudConnect in PeakView software for further evaluation.

RESULTS

Creating ion libraries in the OneOmics suite for use in DIA-NN software

High-pH fractions of K562 cell lysate digest were analyzed using Zeno DDA and processed using the Ion Library app in the OneOmics suite. The Ion Library app is an extension of ProteinPilot app and uses both the Paragon Algorithm and Pro Group Algorithm to infer peptides and proteins from DDA data. The app creates ion libraries in the *.txt file format for DIA data processing, which are compatible with DIA-NN software. Additionally, *.groupexport files are created that can be explored in the ProteinPilot app to assess overall protein and peptide identifications and underlying data quality. The final K562 digest library contained 8,373 proteins at <1% global FDR and 211,150 peptides at <1% global FDR (Figure 2). For each spectral library created in OneOmics suite, a record of analysis settings is saved to help users reproduce their processing results.

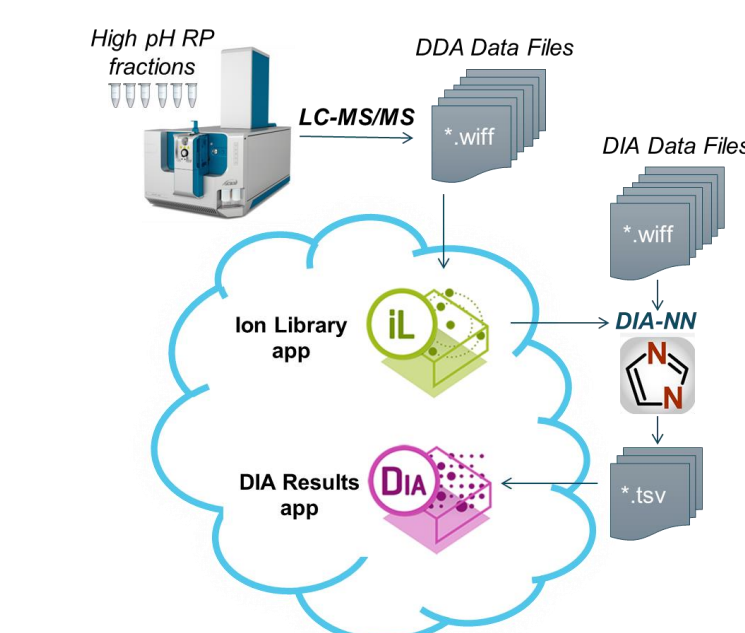


Figure 1. Overview of processing SWATH DIA data using DIA-NN software and the cloud-based OneOmics suite. High-pH fractions of K562 digest were analyzed by microflow chromatography and Zeno DDA on the ZenoTOF 7600 system. The resulting data files were uploaded to the OneOmics suite for processing in the Ion Library app. An ion library was generated and used to process SWATH DIA data in DIA-NN software. Output results files from DIA-NN software were uploaded to the DIA Results app to browse protein and peptide identifications.

Protein Identifications

Proteins identified at critical false discovery rates

	Number of Proteins Detected		
Critical FDR	Local FDR	Global FDR	Global FDR from Fit
1%	7947	8370	8373
5%	8069	8727	8735
10%	8138	9118	9118

Peptide Identifications

Peptides identified at critical false discovery rates

	Number of Peptides Detected		
Critical FDR	Local FDR	Global FDR	Global FDR from Fit
1%	174842	211341	211150
5%	195490	239073	239280
10%	204944	259984	259667

Figure 2. Protein and peptide identifications at critical FDR thresholds for a K562 digest ion library created using the Ion Library app. High-pH fractions of K562 digest analyzed using Zeno DDA were searched in the Ion Library app in OneOmics suite to generate an ion library for DIA-NN software processing. The final library contained 8,373 protein groups and 211,150 peptides at 1% global FDR from fit.

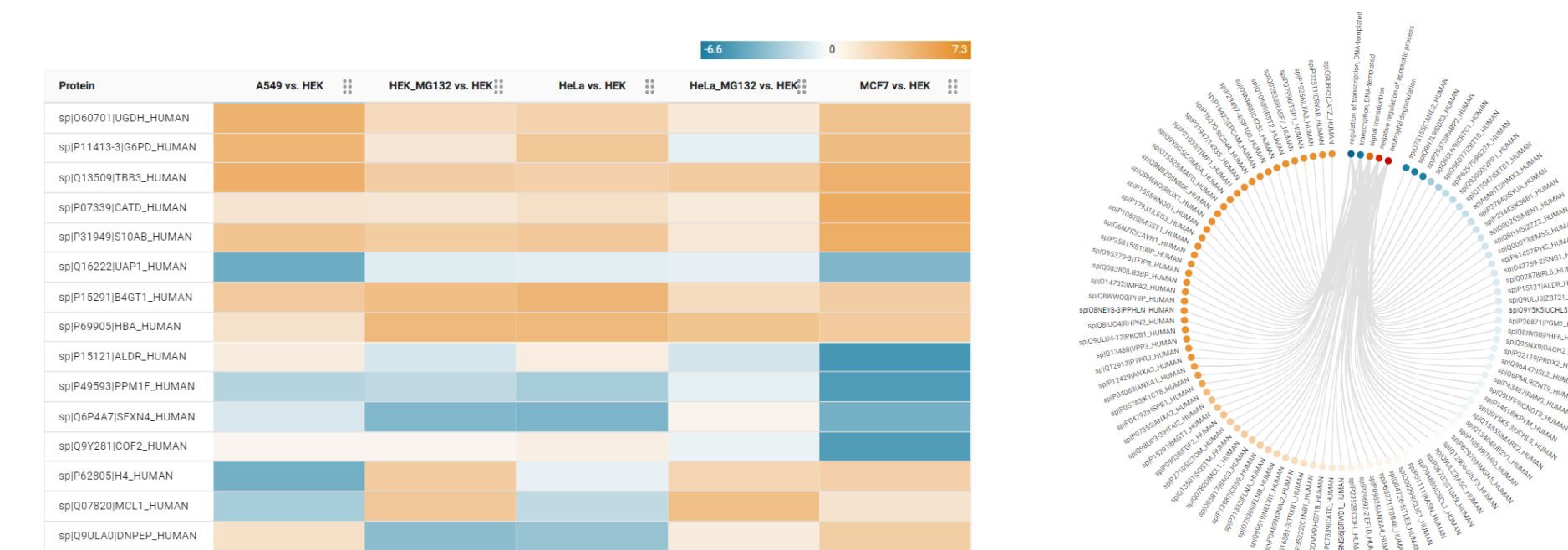


Figure 3. Analysis of differential proteins in the DIA Results app. (Left) A panel of 6 cell lines was evaluated using SWATH DIA on the TripleTOF 6600+ system and DIA-NN software. The DIA results application enables visualization of differential proteins across the samples relative to a specified control in heat map form, with degree of fold change indicated by the depth of color shown (orange: positive fold change, blue: negative fold change). (Right) Differential proteins can be mapped to ontologies corresponding to biological processes, cellular components, and molecular functions. Here, MG132-treated HEK cells were compared to untreated HEK cells, with differential proteins linked to biological processes.

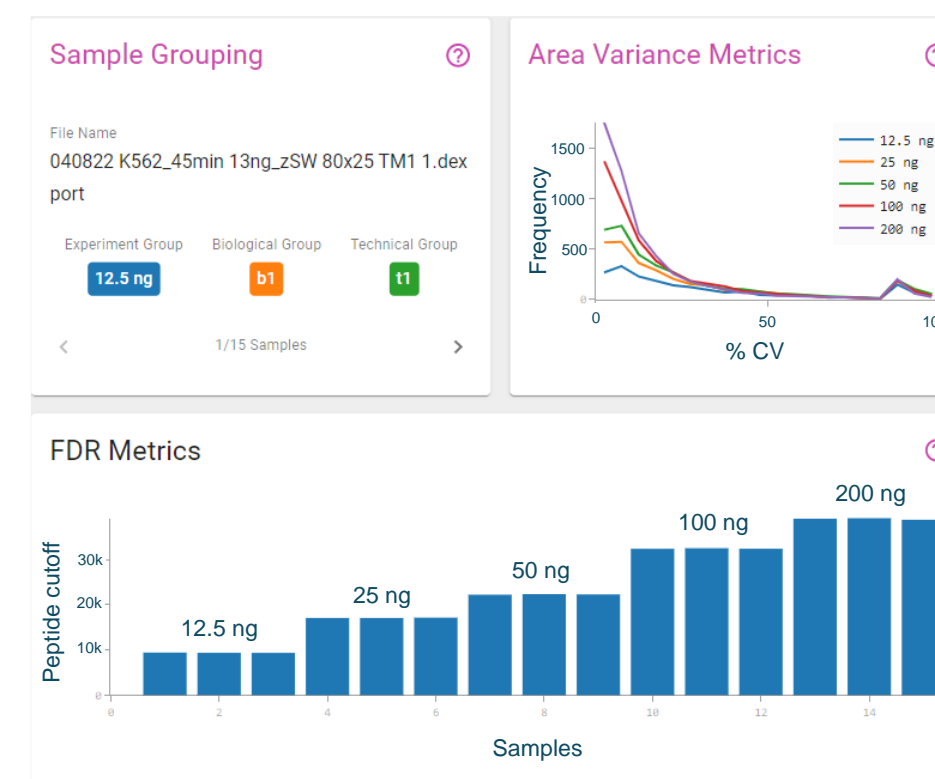


Figure 4. Analysis of Zeno SWATH DIA data using the DIA Results app in the OneOmics suite. Zeno SWATH DIA data collected using a 45-minute microflow LC gradient and 5 µL/min flow rate were processed in DIA-NN software. Results were imported into the DIA Results app for further evaluation. The application generates a series of visuals to assess peptide identifications at a 1% FDR cutoff for each replicate (FDR metrics plot) and illustrates the frequency of peptide identifications at various %CV values (area variance plot). Here, expected results were achieved for loading amounts of K562 digest ranging from 12.5–200 ng for Zeno SWATH DIA results acquired in triplicate. The application also saves a record of the sample grouping used for each experiment to enable reproducible data processing (sample grouping plot).

Evaluating SWATH DIA results in the OneOmics suite

A series of K562 loading amounts ranging from 12.5–200 ng was analyzed using Zeno SWATH DIA on the ZenoTOF 7600 system. The resulting files were processed in DIA-NN software using library-based processing. The resulting *.tsv output files were uploaded to OneOmics suite and imported into the DIA Results app to evaluate identifications. The DIA Results app creates *.dexport files upon transforming the *.tsv results output from DIA-NN software. The *.dexport files can then be further evaluated.

In the DIA Results app, peptide identifications at 1% FDR and critical %CV thresholds are visualized for a quick data quality assessment. In the data set explored here, each K562 load was analyzed in triplicate and evaluated for protein and peptide identifications (Figure 3). The FDR Metrics plot, which indicates peptides passing a 1% FDR threshold in each sample analyzed, illustrated consistent yields within each set of replicates. The highest loading amount (200 ng) yielded the greatest amount of peptides passing a 1% FDR threshold and exhibited the highest frequency of proteins with %CV <20% in the area variance metrics plot. For each experiment, the application also records metadata used during data processing for future reference. An overview of proteins and peptides reported is provided in the sample grouping summary.

Library-free search results viewing in the OneOmics suite

In addition to supporting library-based processing, DIA-NN software also supports library-free searching of DIA data. For library-free processing, FASTA sequences are digested in silico using user-specified settings and the resulting output is used for analyzing SWATH DIA data^{8,9}. Results files from library-free searching created upon using DIA-NN software can be imported into OneOmics suite for statistical analysis and visualization using the same workflow as described here for library-based DIA-NN software search results.

Exploring differential proteins in the DIA Results app

The DIA Results app enables exploration of differential proteins and their associated fold-changes in the form of volcano plots, heat maps and ontologies. Here, a SWATH DIA data set from the TripleTOF 6600+ system consisting of 6 different cell lines was processed in DIA-NN software. Results were imported into OneOmics suite to browse protein fold-changes relative to a specified control. Heat maps enabled rapid visualization of differential proteins and their associated fold changes (Figure 4) and can be custom-filtered by fold-change confidence and reproducibility metrics. Ontology plots are also provided to correlate differential proteins to biological process, molecular functions, and cellular components. The dot beside each ontology term illustrates the direction and ratio of the fold change for the ontology term as determined from the associated proteins. Additionally, PCA scores plots are provided to illustrate sample clustering pre- and post-normalization, along with metrics illustrating ion library coverage. The visualizations provided in the DIA Results app facilitate interpretation of biological studies processed using DIA-NN software.

CONCLUSIONS

Here, a data processing workflow using DDA and SWATH DIA data is demonstrated using DIA-NN software and OneOmics suite. Zeno DDA data were processed in the Ion Library app to generate spectral libraries. The spectral libraries were then downloaded and used in DIA-NN software to process Zeno SWATH DIA data from the ZenoTOF 7600 system. Upon import into the DIA Results app in OneOmics suite, critical protein and peptide identifications were evaluated. The demonstrated workflow also works for SWATH DIA data collected on SCIEX X500 series systems and TripleTOF systems, providing a cloud processing pipeline for all SCIEX accurate mass spectrometers. Using the tools available in the OneOmics suite, this workflow could be extended for the exploration of differential proteins and putative biomarkers by comparing differential proteins across biological samples.

REFERENCES

- Fast-track proteomics data processing with the OneOmics suite SCIEX technical note, RUO-MKT-02-6969-B.
- Uploading and using transcriptomics data in the OneOmics suite. SCIEX community post, RUO-MKT-18-12201-A
- Rapidly advance quantitative proteomics with a high-throughput SWATH acquisition solution. SCIEX technical note, RUO-MKT-02-9710-A.
- Demichev, V et al. (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nature Methods 17, 41–44.
- Architecture and security overview. SCIEX technical note, RUO-MKT-04-6770-B
- Large-scale protein identification using microflow chromatography on the ZenoTOF 7600 system SCIEX technical note, RUO-MKT-02-14415-A.
- Processing ZenoTOF 7600 system data with DIA-NN software, SCIEX community post, RUO-MKT-18-14611-A.
- Going library-free for protein identification using Zeno SWATH DIA and in silico-generated spectral libraries. SCIEX technical note, RUO-MKT-02-14675-A.
- Creating a library from a FASTA file for library-free data analysis SCIEX community post RUO-MKT-18-14611-A.

TRADEMARKS/LICENSING

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 www.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 (see www.sciex.com/trademarks). © 2023 DH Tech, Dev. Pte. Ltd. RUO-MKT-10-15608-A.

To receive a copy of this poster:

- Scan the code with your phone camera
- Complete the form

