Biomarkers and Omics

SCIEX
The Power of Precision

Simplifying the use of ion libraries during data processing of SWATH acquisition proteomics data

Ion library merging with auto retention time calibration in OneOmics suite

Matthew Huebsch[1], Adam Lau[1], Stephen Tate[1], Kathleen Lewis[2], Nick Morrice[3], Sara Ahadi[4], Christie Hunter[2]
[1]SCIEX, Canada, [2]SCIEX, USA, [3]SCIEX, UK, [4]Stanford University, USA

As the use of data independent acquisition (DIA) grows in proteomics research, the need for improved data processing workflows increases. The most common DIA data processing workflow is to use spectral ion libraries to drive targeted extraction of peptide / fragment peak areas from the data, using the m/z and retention time information contained in the library. Increasing the size and quality of the ion library has been shown to increase the number of proteins reliably detected and quantified from a dataset[1]. Retention time (RT) correlation between ion library and the dataset is another key factor that determines quality of data extraction. Currently, the user either doses in a standard peptide mix or manually selects endogenous peptides to be used as retention time markers between the data and the ion library.
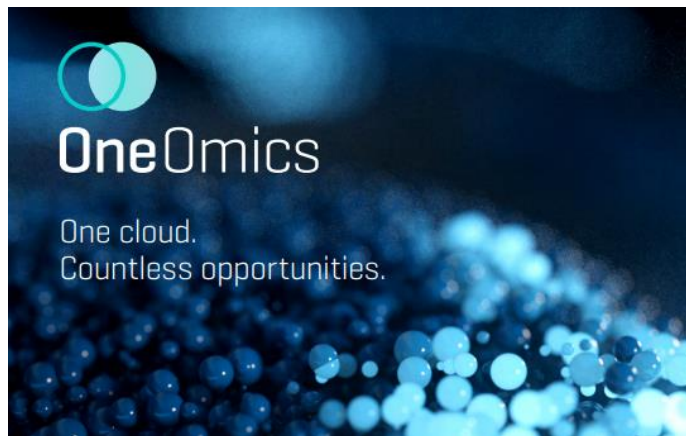
Simplifying how ion libraries are used during DIA / SWATH acquisition data extraction was explored in this work. Two algorithms were developed to simplify data processing within the OneOmics suite cloud processing pipeline, the Library Merging algorithm and the AutoRT Calibration algorithm.

Streamlining the OneOmics suite processing pipeline

- The OneOmics suite for cloud processing of SWATH acquisition data significantly accelerates time to answers for global protein quantitation studies
- The ability to easily combine and expand libraries will enable the extraction of more peptides and proteins from a proteomics sample using SWATH acquisition
  - Retention time alignment during the expansion of libraries is critical to maintain library quality, to allow use of narrow retention time windows during data processing
- During extraction of SWATH acquisition data, it is essential to align the retention times of the libraries with the retention time frame that exists in the data files being processed. An algorithm for retention time calibration to automatically perform this step significantly simplifies sample preparation (no need for RT standards) and data processing.
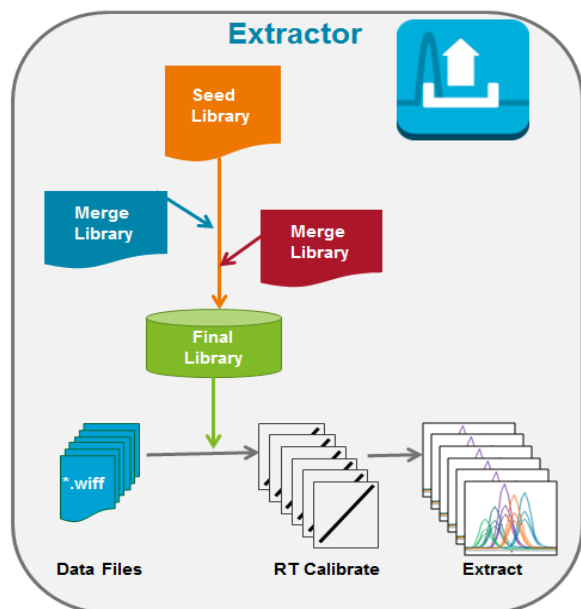

Figure 1. Automated algorithms for simplifying the SWATH acquisition data processing pipeline in the OneOmics suite. Built into the Extractor application in OneOmics suite, the Library Merging algorithm and the AutoRT Calibration algorithm significantly simplifies the DIA processing pipeline. First, new libraries are merged into a seed library using a non-linear RT alignment strategy, then this library is used for SWATH acquisition data extraction by automatically selecting good intensity endogenous peptides across the time range and using these to match the library retention time frame to that of each datafile.

Methods

Sample Preparation: Digested cell lysates were obtained from various sources for testing, including human cell lysates, mouse cell lysates and human peripheral blood mononuclear cell (PBMC) lysates.

Chromatography: Separation of the digests were performed on a NanoLC 425 system (SCIEX) operating in trap elute mode mode at either nanoflow or microflow rates. A 0.3 x 150 cm column and a 0.35 x 10 mm trap were used for the microflow separations. A 75 μm x 150 mm column and a 0.35 x 0.5 mm trap were used for the nanoflow separations. All traps and columns used were packed with both ChromXP™ C18CL, 5 μm, 120 Å phase.

Mass spectrometry: MS analyses were performed using either data dependent acquisition (DDA) or SWATH acquisition on a TripleTOF 6600 system. Either the NanoSpray source or the Turbo V source with a 25 μm I.D. hybrid electrodes was used for ionization (SCIEX). Variable Q1 window SWATH acquisition methods (100 windows)[2] were built using high sensitivity MS/MS mode with Analyst TF software 1.7.1.
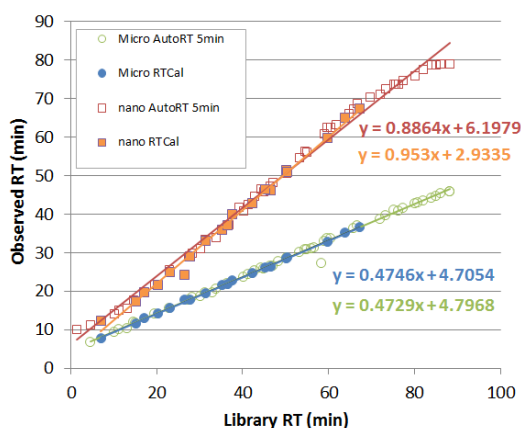
$$y = 0.8864x + 6.1979$$
$$y = 0.953x + 2.9335$$
$$y = 0.4746x + 4.7054$$
$$y = 0.4729x + 4.7968$$

	AutoRT		RT Cal		Difference vs RT Cal	
	# of Peptides	# of Proteins	# of Peptides	# of Proteins	% Peptides	% Proteins
Nano	11871	3373	13035	3486	-8.9%	-3.2%
Micro	10555	3036	10771	3046	-2.0%	-0.3%

Figure 3. Using a microflow generated library to process nanoflow and microflow datasets. Using the standard library from the SWATH Performance kit[5], a nanoflow and a microflow SWATH acquisition dataset were analyzed using 5 minute retention time windows. For the microflow dataset, the AutoRT performed very similarly to the RT Cal with PepCalMix (top, blue and green circles) and provided very similar # of peptides/proteins quantified (bottom). For the nanoflow dataset, there was a slight drop in quantified peptides due to a small difference in gradient profile between the microflow library generation and the nanoflow SWATH acquisitions (red and orange squares).
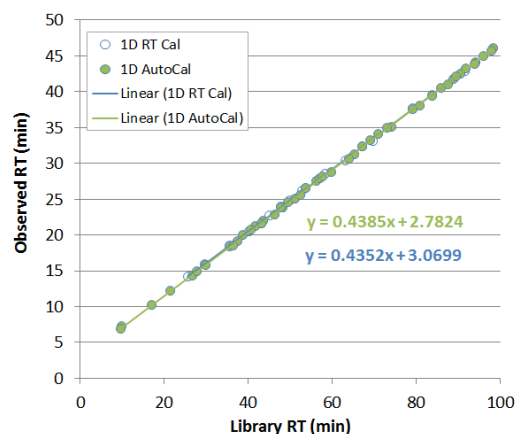
$$y = 0.4385x + 2.7824$$
$$y = 0.4352x + 3.0699$$

	AutoRT		RT Cal		Difference vs RT Cal	
	# of Peptides	# of Proteins	# of Peptides	# of Proteins	% Peptides	% Proteins
Bio 1	4846	1518	4861	1539	-0.3%	-1.4%
Bio 2	4918	1550	4975	1572	-1.1%	-1.4%
Bio 3	4891	1537	4834	1533	1.2%	0.3%
Bio 4	4689	1500	4700	1495	-0.2%	0.3%

Figure 2. Comparing auto retention time calibration to using retention calibration standards. Using a library generated from LC-MS/MS DDA data on a mouse cell line (1D library), 4 mouse samples were analyzed. (Top) AutoRT (green fill) was compared to RT calibration using the peptide standards (PepCalMix, blue circle) and very similar calibration curves were observed. When the # of proteins and peptides quantified were compared (<1%FDR and < 20% CV), the numbers were within 1% indicating very similar calibrations were achieved (Bottom).

Data processing: DDA data was processed with ProteinPilot Software and the group file was used as the spectral ion library. Library files and SWATH acquisition data were uploaded to the cloud using CloudConnect and data were processed using OneOmics suite (Figure 1).

Two algorithms within the OneOmics suite cloud processing pipeline were tested; the Library Merging algorithm and the AutoRT Calibration algorithm. During library merging, ProteinPilot software group files are combined by selecting the largest file as seed library, then peptides from smaller libraries are merged in using a non-linear calibration strategy[3]. New peptides are added to existing proteins and new proteins are added if not present in seed library. During SWATH acquisition processing, endogenous peptides are automatically selected across time range and the best peptides are chosen based on precursor intensity and ID confidence[4]. Best scoring peak groups are used for RT calibration. All comparisons of number of proteins/peptides quantified were done using <1% FDR and <20% CV filters.

Validating the automatic retention time calibration algorithm

To evaluate the quality of the retention time alignment during data processing using AutoRT Calibration, results were compared one the same SWATH acquisition data set using both AutoRT and the typical RT calibration process using a set of spiked standards peptides (RT Cal). In 33/48 validation tests, AutoRT Calibration approach found similar or more proteins quantified than the standard approach and similar linear fit equations (data not shown). The algorithm was then applied to a series of additional biological examples. In the simplest case, a mouse 1D library was generated and used to process the replicates on a set of four mouse cell line samples; both IDA and SWATH acquisition data were collected with the same gradients. The AutoRT algorithm provided a very similar retention time calibration line to the standard synthetic peptide approach and resulted in very similar number of proteins and peptides quantified (Figure 2).

Next, the use of libraries generated at different flow rates was explored. A large human library collected using microflow on 15 high pH fractions was used to process SWATH acquisition replicates collected at both nano and microflow rates (Figure 3, SWATH Performance Kit Library[5]). The RT calibration lines for microflow (RT Cal - blue and AutoCal - green) were very similar and resulted in very similar numbers of proteins and peptides quantified. There was a slight difference in RT calibration when using the microflow library with nanoflow data, most likely because small changes in gradient profiles, seen by the light curvature in the red points at the higher organic end of gradient.
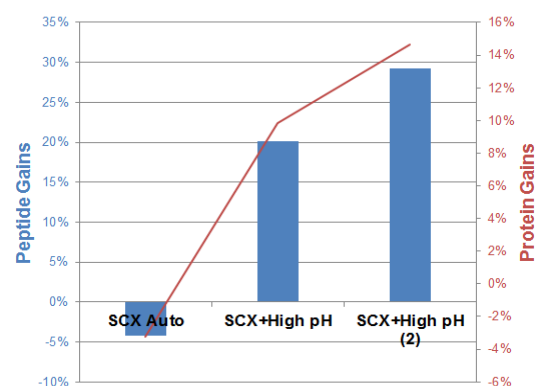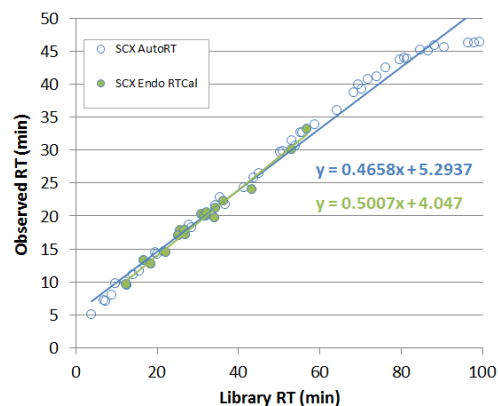

$$y = 0.4658x + 5.2937$$
$$y = 0.5007x + 4.047$$

Figure 4. Multiple orthogonal fractionation experiments to generate a larger PBMC ion library. Protein ID results (ProteinPilot software group files) obtained from processing individual fractionation experiments were sequentially combined and then used to process a set of SWATH acquisition replicates. (Top) A single SCX library was compared using AutoRT and RT calibration with manually selected endogenous peptides, very similar linear fits were obtained. (Bottom) Additional fractionation libraries were then merged in and used to process the same SWATH acquisition data, increased number of peptides and proteins were quantified as the merged library grew.

This resulted in a small drop in the number of peptides quantified using the AutoRT approach. This highlights the importance of using similar linearity of gradient profiles across both library generation and SWATH acquisition experiments (gradient duration can change but profile should be similar).

Integrated library merging with autoRT calibration

To explore the effect of the merging of libraries, a series of libraries generated from 2D-LC fractionation runs performed on PBMC digests were combined and used to extract SWATH acquisition data replicates on PBMCs digests. First, the use of a single SCX library was compared using AutoRT and RT calibration with manually selected endogenous peptides. Very similar linear fits were obtained with both techniques and a similar numbers of peptide and proteins were quantified (within
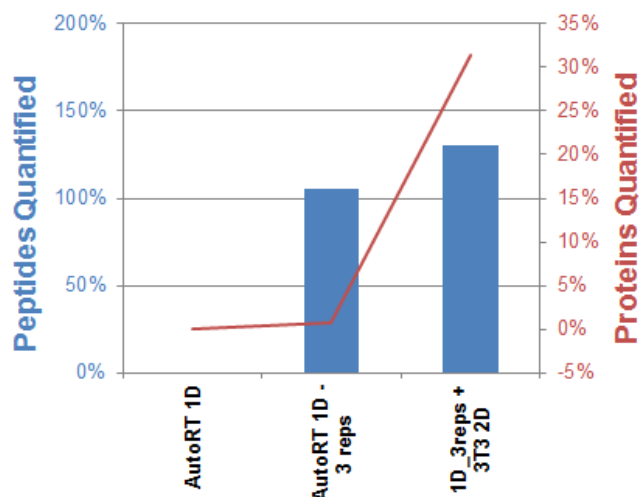

Figure 5. Merging libraries from mouse cell lines types. SWATH acquisition data was collected on a mouse cell line, then processed with a 1D library generated on the same sample with either a single DDA acquisition or 3 DDA acquisitions. Next a 2D LC-MS/MS library generated from mouse 3T3 cells was merged in and SWATH acquisition data was reprocessed. All used AutoRT calibration.

4%, Figure 4, top). Next the libraries were sequentially combined and each was used to process the same PBMC SWATH acquisition data; results were compared to the results obtained using the first SCX 2D library and RT calibration with endogenous peptides. Combining the high pH library with the initial SCX library provided a gain of 20% peptides and 8% proteins quantified, as new peptides were added to the library from the orthogonal fractionation experiment. A second high pH library was also merged in, and provided additional but smaller gains. Note these fractionation libraries were generated using manual spin columns so less fractionation was achieved.

Next, a series of SWATH acquisition experiments on Mouse cell lines were studied. Here a library generated from a 1D data dependent acquisition was used to process the SWATH acquisition data and the AutoRT calibration algorithm was used, this was the benchmark (Figure 5). Next, a set of three 1D DDA runs were processed into a single library and used to extract the same SWATH acquisition data. Significant gains in peptides quantified was observed (105%) as the multiple DDA runs provided more peptide coverage for the higher abundant proteins, but minimal increase in proteins quantified. However, a larger 2D library generated from mouse 3T3 cells was then merged into the 1D library and again used to process the same SWATH acquisition data. In this case the expanded library did provide additional proteins that were quantifiable out of the SWATH acquisition data (31% over 1D library). This resulted in ~2000 proteins quantified from the mouse cell lines with ~9900 peptides quantified.

Conclusions

Two algorithms for merging ion libraries and performing retention time calibration have been developed to further streamline the SWATH acquisition data processing pipeline in OneOmics suite.

- The performance of the AutoRT Calibration algorithm was demonstrated to be similar in performance to the RT Calibration process using manually selected peptides or dosed synthetic peptides, providing similar numbers of peptides quantified, while providing a significant improvement in ease of use
  - Even translating microflow libraries for use with nanoflow SWATH acquisition data worked well (Figure 3), providing similar linear equations after retention time calibration

- Merging of relevant 1D and 2D LC-MS/MS ion libraries provides gains in proteins and peptides quantified, especially when additional libraries were generated using orthogonal 2D separations to better protein cover the similar sample type.
- Ability to visualize the retention time calibration lines highlights the importance of using similar, linear LC gradients are used when planning on sharing and merging libraries.

References

1. Extending Depth of Coverage with SWATH acquisition Using Deeper Ion Libraries, SCIEX Technical note RUO-MKT-02-3247-A.
2. Improved Data Quality Using Variable Q1 Window Widths in SWATH acquisition, SCIEX Technical note RUO-MKT-02-2879-B.
3. Merging of Libraries in OneOmics suite, SCIEX community post RUO-MKT-18-6401-A.
4. Calibration of Retention Time in OneOmics suite, SCIEX community post RUO-MKT-18-6401-A.
5. SWATH Performance Kit, SCIEX product page.

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