TIMS DIA-NN: CCS-Aware DIA Data Analysis

Robin Park[1]; Tomasz Siwarga[1], Tharan Srikumar[1]; Jonathan Krieger[1]; Qin Fu[2]; Patrick Garett[3]; Manasa Vegesna[2], Niveda Sundararaman[2], Marc-Antoine Beauvais[1];Sven Brehmer[1]; Chistopher Adams[1]; Dennis Trede[1]; Gary Kruppa[1]; John Yates[3]; Jennifer E. Van Eyk[2]; Rohan Thakur[1], Bruker Daltonik GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany 2Bruker Daltonics Inc., 40 Manning Road, Manning Park, Billerica, MA 01821, USA

Introduction to TIMS DIA-NN

Data-independent acquisition (DIA) has become a vital technology with its reproducibility and deep proteomics coverage. However, DIA data analysis is still challenging because of large data size, fast scan speeds, deconvolution of complex spectra, and calculation of accurate quantitative values using a spectral library. Recently, DIA-NN [ref] has been a breakthrough in DIA data analysis boosting the number of protein identifications by using a neural network (NN). We have re-designed the open-source version of DIA-NN and developed algorithms utilizing collision cross-section (CCS) to improve the identification and quantification in the DIA data. We also developed a spectral library generation tool and integrated it into PaSER with CCS values for the downstream data analysis. Collectively these new features significantly advance the processing of dia-PASEF data and we call this new approach TIMS DIA-NN.

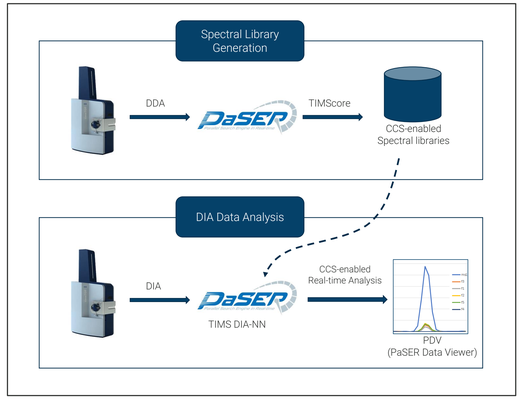

Fig. 1 Spectral library generation. PaSER, real-time search platform, simplifies building CCS-enabled spectral libraries from DDA and running DIA analysis

Methods

We have developed a spectral library generation tool that collects DDA data from PaSER real-time database search platform. The GPU-powered real-time search engine generates unfiltered results in sqt file format. To identify confident protein and peptide list, it generates CCS values from the CCS prediction model built from neural network and calculate TIMScore to classify target and decoy PSMs before generating the final results. The spectral library tool collects the search results and generates the libraries including CCS values. When it builds libraries, it calculates mean and standard deviation of ion mobility values for each precursor. TIMS DIA-NN calculates ion mobility-based scores for both target and decoy precursors and adds them to existing features for deep learning.

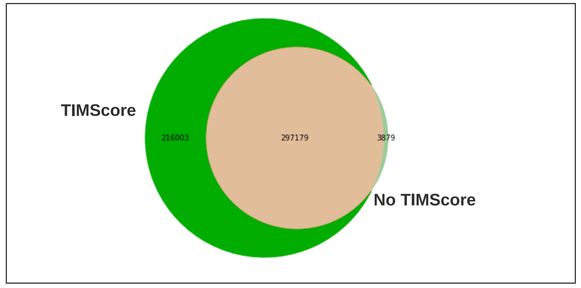

Fig. 2 Number of precursors in the spectral libraries. We compared two spectral libraries from K562 fractionated DDA dataset (35 min gradient) with and without TIMScore.

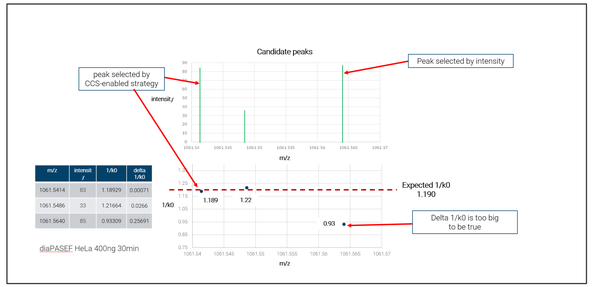

Fig. 3 TIMS DIA-NN utilizes CCS to find correct peaks. It is important to select correct peaks not only improve identifications, but also measure quantitative values accurately.

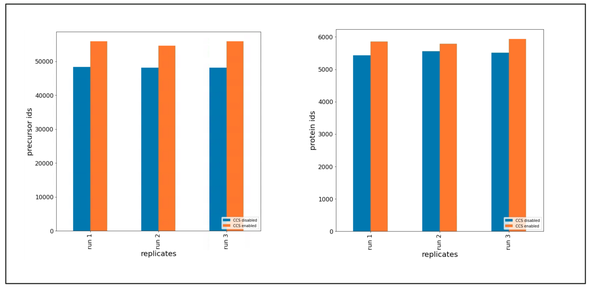

Fig. 4. We ran TIMS DIA-NN on HeLa triplicates with and without CCS algorithms. CCS enabled results show more identifications on both precursor and protein levels

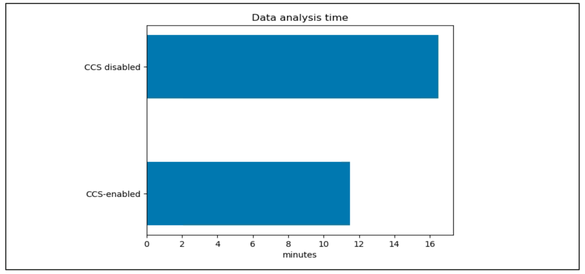

Fig. 5 We compared the performance of DIA analyses with and without CCS algorithms. Interestingly, CCS-enabled run shows better performance. The reason is that CCS-enabled run selects the correct peaks better so it can calibrate faster, reducing the number of iterations.

Plasma samples

Gender pooled plasma (K2EDTA) was purchased from Bioreclamation. Plasma protein denaturation, reduction, alkylation, digestion were performed using the Beckman i7 automated workstation (Beckman Coulter). Briefly, plasma proteins were denatured with L 2,2,2-trifluoroethanol (TFE, Sigma), reduced by dithiothreitol (Sigma) and denatured for 1 h at 60°C. Samples were then alkylated for 30 min at 25°C in the dark with iodoacetamide (Sigma) and digested by Trypsin for 4 hours at 42°C. Digestion reactions were quenched with formic acid. Desalting was carried out using a positive pressure apparatus (Amplius Positive Pressure ALP, Beckman Coulter).

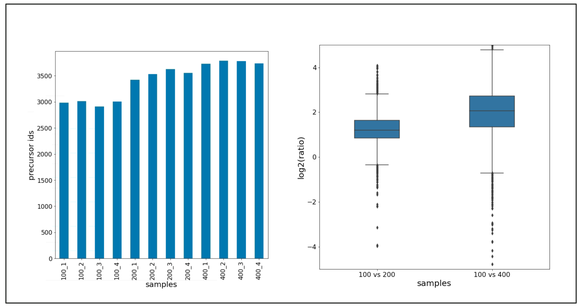

Fig. 6. Numbers of identified precursors for different sample loadings (left). Log ratios of 100 vs 200 and 100 vs 400 samples.

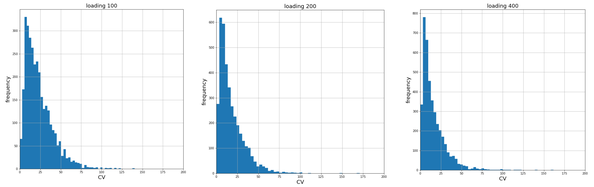

Fig. 7 CVs of three different sample loadings

Conclusion

- PaSER, a GPU powered real-time search platform can build aware spectral libraries
- TIMS DIA-NN, CCS-aware DIA analysis software, is integrated into PaSER
- We re-designed a neural network and produced more than 100 features to boost the number of protein identifications and performance

Technology