Automated Workflows for dia-PASEF Data using TIMS DIA-NN on the PaSER Platform

US HUPO 2022, Poster P08.01

Christopher Adams[1], Robin Park[1], Sven Brehmer[2], Tharan Srikumar[3], Nagarjuna Nagaraj[2], Jonathan Krieger[3]
[1]Bruker Scientific LLC, San Jose, CA [2]Bruker Daltonics GmbH & Co.KG, Bremen, Germany [3]Bruker Ltd., Milton, ON

Introduction

Parallel search engine in real-time or PaSER was developed to take advantage of GPU-powered database search. The GPU-powered ProLuCID-4D algorithm can process the large number of MSMS spectra generated by the PASEF process on the timsTOF platform, while utilizing all four dimensions – retention time, CCS value, m/z and fragment spectra – to increase the confidence in each identification. PaSER has now been extended to take advantage CCS values through the use of TIMScore which results in a dramatic increase in the identification of precursors and protein identifications. The ability to use TIMScore with DDA data is seamlessly integrated on PaSER for spectral library generation, resulting in more robust spectral libraries. These PaSER built libraries can then be used to search dia-PASEF data in real-time on the PaSER box using the first vendor integrated version of DIA-NN (TIMS DIA-NN). Thus, the PaSER platform offers a full solution for DIA and DDA analysis in real time, allowing access to meaningful data as soon as acquisition ends.

TIMScore Leverages CCS Values to increase identification depth

TIMScore is available as an update on all PaSER software, and uses machine learning to predict peptide CCS values, and functions to define the deviation between experimental and predicted CCS values.

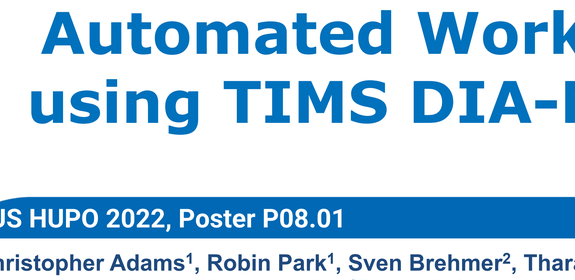

Figure 1 – Accuracy of TIMScore model was tested against a previously unseen dataset. Correlation between predicted and observed CCS values were calculated as (A) 95% for tryptic peptides and (B) 92% for phosphorylated tryptic peptides, respectively.

The benefit of TIMScore can be realized during False Discovery Rate (FDR) estimation. In a non CCS enabled algorithm, only two dimensions can be utilized to estimate the FDR rate, and so a discriminate line is fit to a 1% error. With TIMScore, and the extra CCS dimension, the peptide-candidates can be vectorized in 3-dimensions allowing a discriminate contoured plane to be applied to achieve the same 1% error.

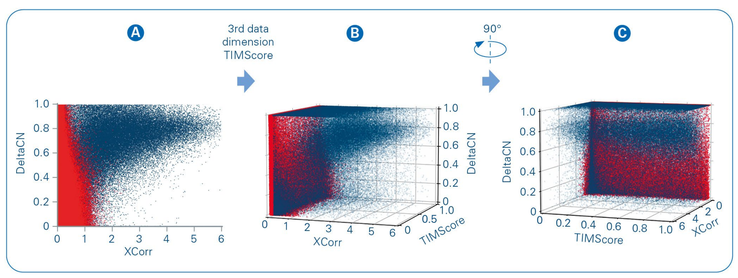

Figure 2 – 2D (A) representation of DeltaCN and Xcorr compared to 3D (B,C) representations when the TIMSscore dimension is added.

Even more telling is the comparison of a real dataset with and without TIMScore. The published dataset from the Ishihama lab (PXD019746), in which a series of mono-phosphopeptide and their corresponding unphosphorylated peptides was reprocessed with the PaSER 2022c platform. A significant increase in precursor, peptide and protein identification is observed.

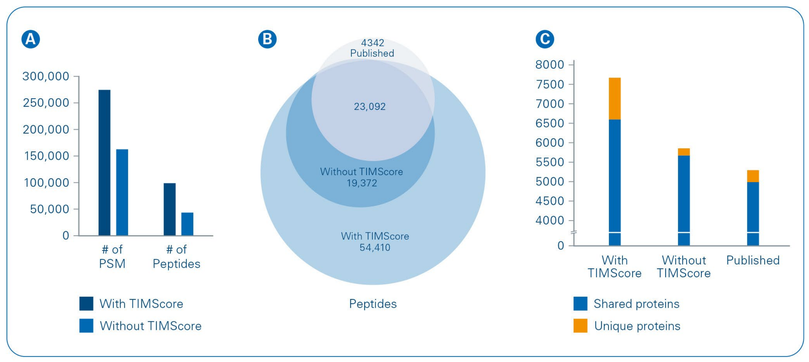

Figure 3 – (A) Bar graph of the number of identified peptide-spectrum-matches with and without TIMScore. (B) Venn diagrams displaying all peptide sequences and (C) protein groups identified. All results are from the published dataset PXD019746 as processed and presented without TIMScore, with TIMScore and as in the published research article.

Using TIMScore enabled spectral libraries for integrated TIMS DIA-NN dia-PASEF Analysis

PaSER 2022c fully integrates spectral library generation with a powerful CCS-Enabled vendor integrated version of DIA-NN called TIMS DIA-NN. This integrated solution allows the real-time interrogation of dia-PASEF data in an environment that fully supports and takes advantage of the IM dimension.

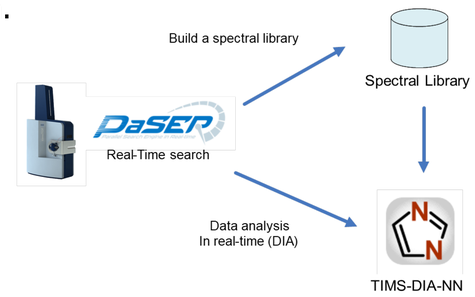

Figure 4: Schematic of workflow for integrated library generation and TIMS-DIA-NN in real-time

Methods

K562 tryptic digests (Promega) were used for benchmark measurements by coupling a nanoElute (Bruker) with n Bruker-25 cm column attached to a trapped ion mobility spectrometry – quadrupole time of flight mass spectrometer (timsTOF Pro 2). A fractionated set of K562 data run in DDA mode was used to generate spectral libraries, and then an optimized dia-PASEF scheme method (35 min gradient) was used to generate dia-PASEF data. DDA data was searched using the ProLucid GPU powerd search engine, and DIA data was searched using TIMS-DIA-NN all within the PaSER environment.

Results

Two spectral libraries were made using the same K562 fractionated dataset, one with TIMScore enabled and one without TIMScore. The addition of TIMScore resulted in an increase of ~216k precursors compare to No TIMScore, at the same 1% FDR cutoff, allowing for the generation of a significantly more robust spectral library for DIA searches.

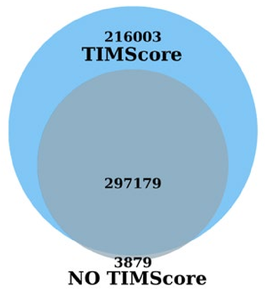

Figure 5 – Venn Diagam illustrating overlap in precursors between spectral libraries built with TIMScore and without.

When using TIMS DIA-NN and a TIMScore enabled library, the true advantage of a CCS-enabled ecosystem is realized. With 200ng of starting material, nearly 9000 protein groups (12,000 precursors) can be identified in a 35minute gradient. Quantification of identified precursors and proteins is accurate, with small CVs.

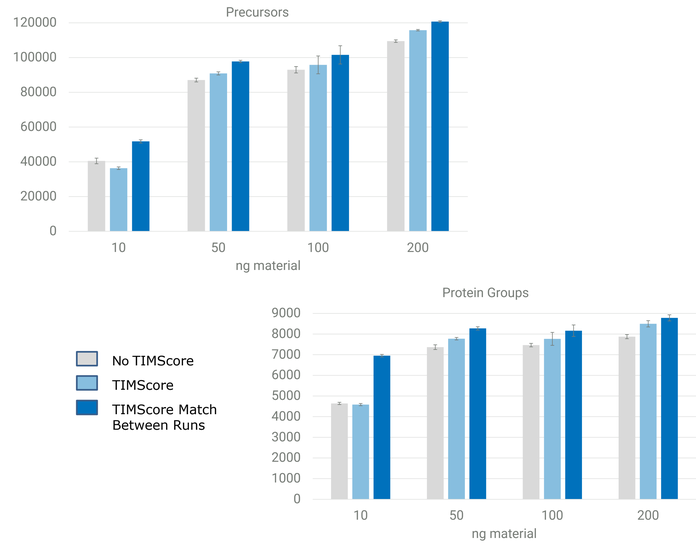

Figure 6 – Bar charts illustrating precursor and protein group numbers of varying amounts of K562 run in dia-PASEF mode, searched with TIMS-DIA-NN

Conclusions

- **TIMScore and TIMS DIA-NN enabled on the PaSER platform provide a real-time CCS-enabled solution for analysis DIA and DDA PASEF data**
- **These tools allow new depths of identification to be realized by taking advantage of the ion mobility dimension on the timsTOF series of instruments**

PaSER/timsTOF Pro/timsTOF SCP