HUPO 2024 – ABSTRACT 453

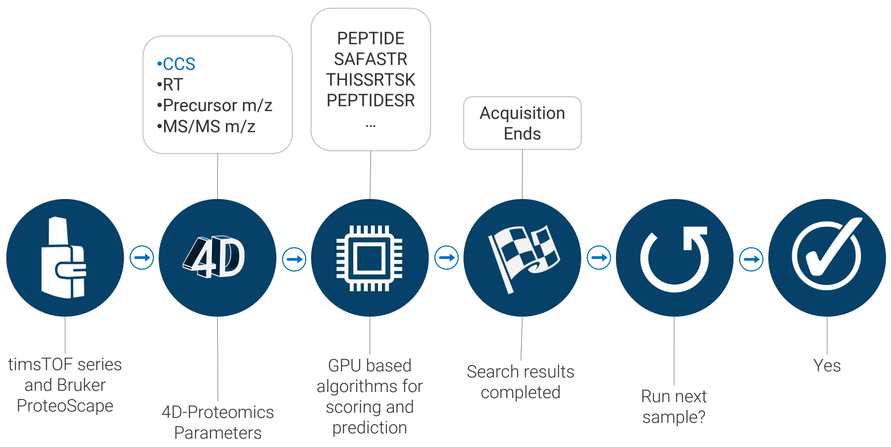
A CCS-centric HLA-specific trained de novo module for precise and accurate real-time immunopeptide identification on the Bruker ProteoScape™ platform

Rui Zhang[1]; Qixin Liu[1]; Dennis Trede[2]; Tharan Srikumar[3]; Jonathan Krieger[3]; Bin Ma[1]; George Rosenberger[4]; Angela Paul[5]

[1]Rapid Novor Inc., Kitchener, ON; [2]Bruker Daltonics GmbH & Co. KG, Bremen, Germany; [3]Bruker Ltd., Milton, ON; [4]Bruker Switzerland AG, Faellanden, Switzerland, [5]Bruker UK Ltd, Coventry United Kingdom

Introduction

Bruker ProteoScape (BPS) has transformed into a comprehensive proteomics data analysis platform that can integrate third-party tools while utilizing the concept of data streaming to realize fully customizable real-time processing workflows. To expand the capabilities of the BPS platform for immunopeptidomics, and other applications, we previously developed and integrated a timsTOF optimized de novo sequencing engine from Rapid Novor Inc., called BPS Novor. To further develop this tool and support the extremely rapidly growing field of timsTOF based immunopeptidomics, we have retrained this module with 1.4 Million PSMs mapping to >150,000 HLA peptides specifically generated from MHCI and MHC II samples (Hoenisch Gravel et al., 2023) .



Methods

Novor was re-trained on a variety of timsTOF acquired data, where ground truth is taken from the ProLuCID database search results filtered to 1% PSM FDR. Previously shown that BPS Novor is highly accurate and precise on a variety of datasets. On amino acid level, at 75% precision, BPS Novor achieved between 40-60%, whereas standard Novor achieved between 25-50%. We have also shown that BPS Novor is extremely fast by evaluating the processing speed across 5 datasets, with an average processing speed of 1338±226 spectra/second.

Here we compare the re-trained BPS Novor (MHC model) across two immunopeptidomics datasets using standardized numbers of computing cores. We focused on two recently published immunopeptidomics datasets (Feola et al., 2021 and Phulphagar et. al., 2023) to show the newly optimized BPS Novor MHC model increases precision and accuracy over the previous versions for immunopeptidomics applications, while retaining the processing speed advantage over other algorithms. All data processing was conducted on a first generation PaSER workstation, consisting of 16 core processor (AMD Epyc 7302P) with access to 128GB RAM. Processing time was computed as the time taken to load MS2 spectral data, score spectra and generate output files.

Fig. 1: Run & Done: The BPS platform allows data to be streamed during acquisition for processing on a separate workstation, allowing for real-time data analysis, so that when acquisition completes, users have data for review.

Results

To demonstrate the performance improvements for immunopeptidomics applications, we compared the sequencing results of BPS Novor with three different scoring models: 1) Pre-training pre-optimization Orbitrap HCD model; 2) timsTOF optimized unspecific digest model; and 3) timsTOF optimized MHC model. We first looked at three samples from Feola et. al, where we saw ~2% improvement in amino acid correctness, ~1% increase in peptide correctness. Both timsTOF optimized models outperformed the pre-training Orbitrap HCD model as excepted. While the larger MHC model also increased the de novo processing time, it was marginal particularly in relation to the LC gradient.

We next looked at set of 10 melanoma derived samples from Phulphagar et. al. A ~6% increase in amino acid and peptide correctness was observed. The MHC model showed on average 8.5% increase in the number of 8-11mer peptide sequences in the dataset. Together, these datasets confirm that the MHC optimized scoring model for BPS Novor provides greater accuracy for immunopeptidomics data.

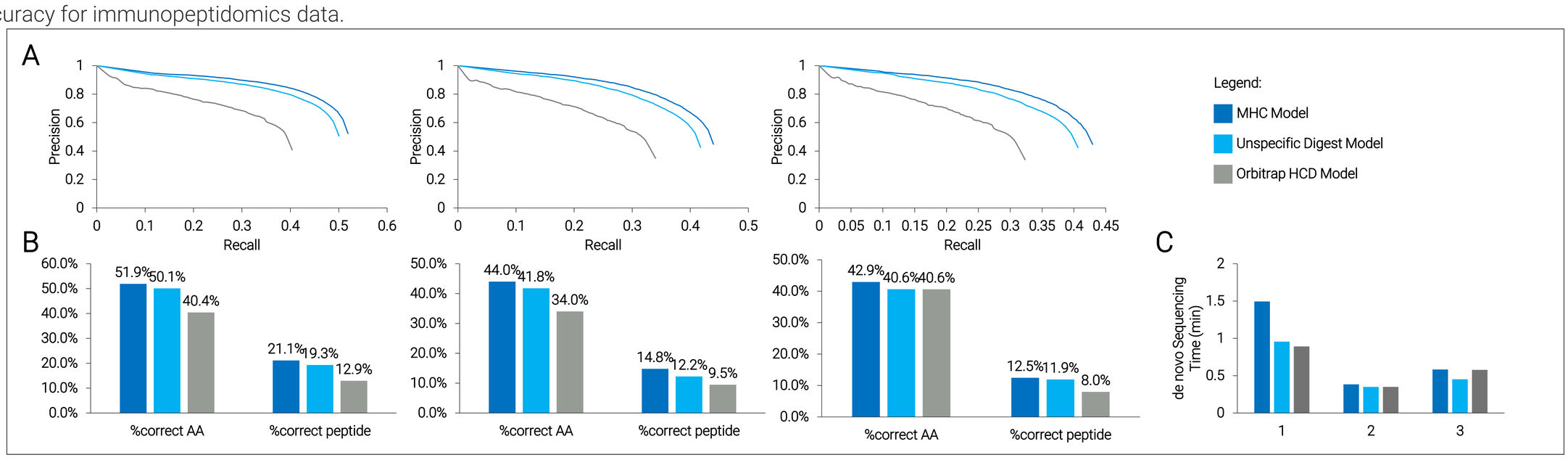


Fig. 2: Amino acid precision-recall graphs (A) for three MHC class 1 eluted samples from PXD022194 (Feola et al., 2021) . The percent of correct amino acids and peptides assigned by each scoring model is also shown (B). (C)The time required to process these dataset are illustrated showing negligible changes due to the scoring models, particular in relation to the LC gradient of 45min.

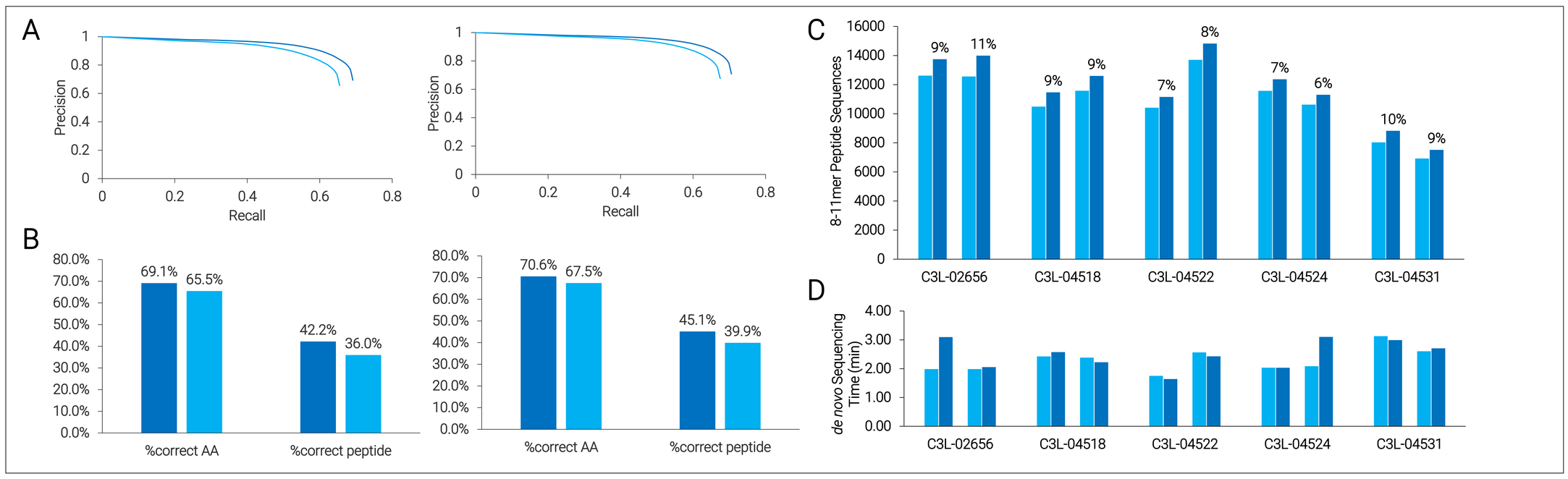


Fig. 3: Evaluation of MHC model on low-input primary melanoma tumor derived HLA-I immunopeptidomes from MSV000091456 (Phulphagar et al., 2023). The 10 samples were reprocessed using BPS's standard unspecific digestion model and the MHC model. (A) Amino acid precision-recall graphs for representative samples along with (B) the percent of correct amino acids and peptides assigned by each scoring model. (C) The number of 8-11mer peptide sequences identified in each of the samples with both models (with percentage gain with the MHC model). (D) The processing time required for each of these 120min dda-PASEF acquisitions for each model.

References:

1) Hoenisch Gravel, N. et al. TOFIMS mass spectrometry-based immunopeptidomics refines tumor antigen identification. Nat Commun 14, 7472 (2023).
2) Phulphagar, K. M. et al. Sensitive, High-Throughput HLA-I and HLA-II Immunopeptidomics Using Parallel Accumulation-Serial Fragmentation Mass Spectrometry. Molecular & Cellular Proteomics 22, (2023).
3) Feola, S. et al. PeptiCHIP: A Microfluidic Platform for Tumor Antigen Landscape Identification. ACS Nano 15, 15992–16010 (2021).

Conclusion

- A fast, accurate and precise peptide de novo sequencing algorithm has been integrated into BPS, providing Run & Done capabilities to additional 4D-Proteomics applications.
- A purposefully optimized MHC scoring model shows noticeable increase in accuracy for immunopeptidomics datasets with minimal change in computation time.

Technology

© 2023 Bruker

For Research Use Only. Not for use in clinical diagnostic procedures.

Innovation with Integrity