TIMSrescore: timsTOF-optimized PSM rescoring boosts identification rates for immunopeptidomics

Dennis Trede[1], Arthur Declercqi[2], Ralf Gabriels[2], Robbin Bouwmeester[2], Jonathan Krieger[3], Tharan Srikumar[3], Sven Degroeve[2] and Lennart Martens[2]

[1]Bruker Daltonik GmbH, Fahrenheitstraße 4, 28359 Bremen, Germany [2]VIB-UGent Center for Medical Biotechnology, Belgium [3]Bruker Ltd., 2800 Highpoint Dr. Suite#206, Milton, ON L9T 6P4, Canada

Introduction:

The immunoproteome, the subset of protein involved in immune response, are key components of various disease mechanisms and recent progress in MS instrumentation and analysis have gained considerable attention for this special application. However, because immunopeptides are non-enzymatically cleaved, the necessary search space dramatically increases, resulting in lower peptide identification rates. It has been demonstrated that the introduction of orthogonal scores comparing predicted vs observed peptide properties can increase both sensitivity and specificity in these scenarios. Here, we introduce TIMSrescore, a timsTOF-optimized algorithm based on MS2Rescore, that provides these benefits for the timsTOF family of instruments.

The peptide fragmentation (MS2PIP) and collisional cross section (IM2Deep) predictors employed in TIMSrescore have been trained using a diverse set of timsTOF dda-PASEF PSMs. The database search engine Sage and MS2Rescore have been extended to support timsTOF data. For evaluation we used 2 datasets. 1) Replicate injections of HeLa protein lysate digested with elastase. 2) The RCC tumor samples from Hoenisch et al., 2023. were also reprocessed with our pipeline. Representing triplicate measurements of HLA class I and II enriched samples.

Sage (Lazear et. al, 2023) and MS2Rescore (Declercq et.al, 2022) were run with default parameters for elastase and immunopeptidomics analyses (8-10 AA). For the whole cell lysate datasets, TIMSrescore increased recovery of peptides by >30%. For the immunoproteomic dataset, a considerable increase of 21.8% could be observed, indicating that rescoring is particularly impactful for large search spaces. Particularly the timsTOF optimizations of the fragmentation predictors were critical, improving median Pearson correlation to 0.88 from 0.53 (standard MS2PIP HCD model).

Methods:

For HeLa studies, HeLA protein lysate was reduced, alkylated and digested with protease elastase at 37 °C for 3h. Peptides were analyzed with a nanoElute 2 (gradient times of 22 min) coupled to a CSI Ultra on a TIMS-TOF Ultra mass spectrometer. Intensity threshold (500), target intensity (20,000) and collision energies (1/k0 0.7 - 20 eV, 1/k0 1.06 - 30 eV, 1/k0 1.1 - 40 eV, 1/k0 1.34 - 40 eV, 1/k0 1.68 - 70 eV) were used based on optimization experiments. Singly charged precursors (600 m/z -1,400 m/z) and multiply charged precursors (300 m/z -800 m/z) were included in precursor selection. Varying TIMS accumulation times in combination with “300 Hz” enabled and adjusted No. of PASEF ramps for short cycle times were tested to yield highest identification results. The raw data were processed via Sage search algorithm followed by MS2 Rescore (TimsRescore Branch, Declercq et.al, 2022) and human Uniprot database; enzyme: unspecific; carbamido-methylation (C, fixed); oxidation (M, variable). Peptides are reported at 1% peptide FDR. For reprocessing of the RCC samples, the Sage-MS2Rescore pipeline was used with the settings as described in the original publication.

Results:

HeLa Elastase Experiments

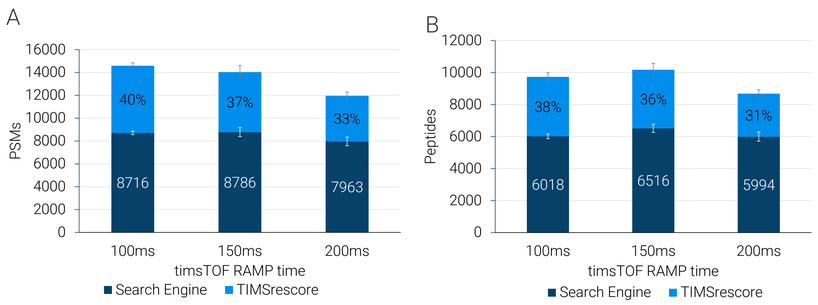

Fig. 1: TIMSrescore increases confidently identified peptides by >30%. 1ng of HeLa cell lysates digested with elastase were analyzed with 22min active LC gradient on timsTOF Ultra with three different TIMS ramp times. The average number of PSMs (A) and peptides (B) are shown with and without rescoring. On average, across all ramp times, rescoring increases PSM identifications by 37% and peptide identifications by 35%.

Reprocessing of RCC tumor samples

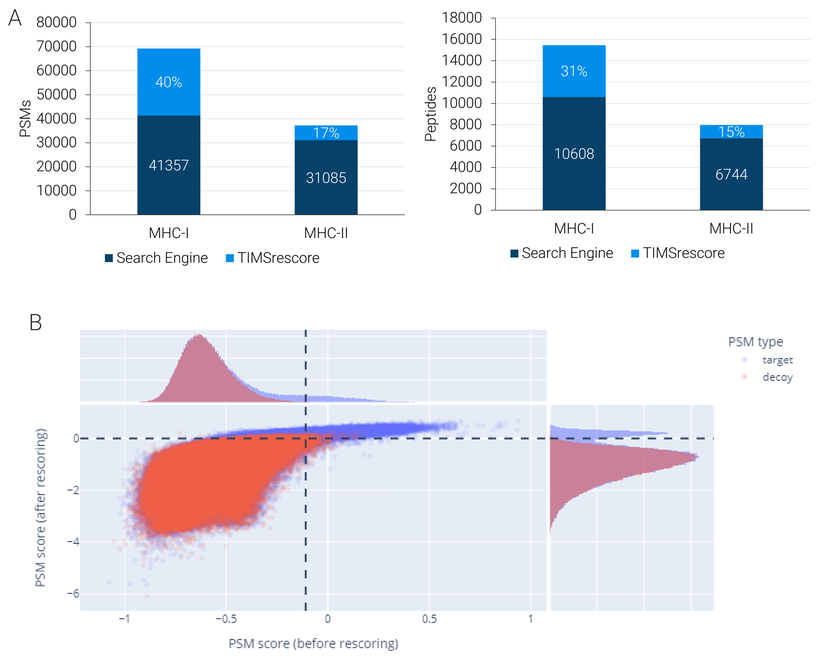

Fig 2: TIMSrescore for Immunopeptidomics. (A) PSM and peptide identifications with and without rescoring for both MHC-I (n=3) and MHC-II datasets (n=3). (B) Scatterplot of Score Comparison for the MHC-I dataset. Target (blue) and decoy (red) PSMs before rescoring are shown on the x-axis and after rescoring are shown on the y-axis. The upper left quadrant are the PSMs only identified after rescoring. (C) False Discovery Rate Comparison. This shows the number of identified target PSMs in function of the FDR threshold. (D)Identification overlap showing PSMs and Peptides removed, retained and gained by the rescoring engine.

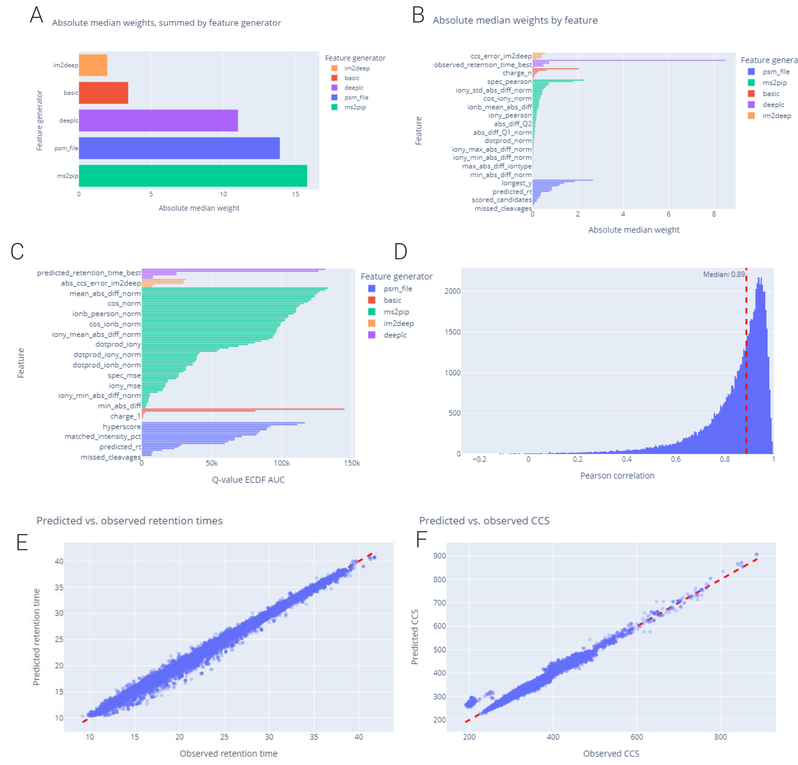

Fig. 3: Feature Usage in TIMSrescoring model for the MHC-I dataset (n=3): (A) Absolute median weights summed by feature generator or (B) broken down by component of each feature generator. Each feature generator contributes to the overall rescoring process of separating target and decoy PSMs. (C) Individual feature performance plot, showing the q-values calculated for each feature, as if that feature were used for rescoring individually. (D) MS^{2}PIP distribution of Pearson coefficients for all target PSMs (E) Deep-LC, and (F) IM2Deep model performance showing predicted vs observed RT and CCS.

References:

Hoenisch Gravel, N. *et al.* TOFIMS mass spectrometry-based immunopeptidomics refines tumor antigen identification. *Nat Commun* **14**, 7472 (2023).

Lazear, M. R. Sage: An Open-Source Tool for Fast Proteomics Searching and Quantification at Scale. *J. Proteome Res.* **22**, 3652–3659 (2023).

Declercq, A. et al. MS2Rescore: Data-Driven Rescoring Dramatically Boosts Immunopeptide Identification Rates. Molecular & Cellular Proteomics 21, (2022).

Conclusion:

- TIMSrescore represents a timsTOF-optimized rescoring approach that can improve recovery of peptide identifications
- While immunopeptidomic samples wildly vary, an increase of 15-40% in PSMs and peptide identified by TIMSrescore can be observed
- TIMSRescore makes use of all available dimensions of timsTOF data including fragmentation, retention time, and the TIMS dimension

Technology