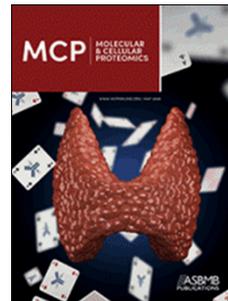


Journal Pre-proof

Benchmarking Software for DDA-PASEF Immunopeptidomics

Yannic Chen, Annica Preikschat, Annette Arnold, Riccardo Pecori, David Gomez-Zepeda, Stefan Tenzer



PII: S1535-9476(25)00591-2

DOI: <https://doi.org/10.1016/j.mcpro.2025.101492>

Reference: MCPRO 101492

To appear in: *Molecular & Cellular Proteomics*

Received Date: 7 July 2025

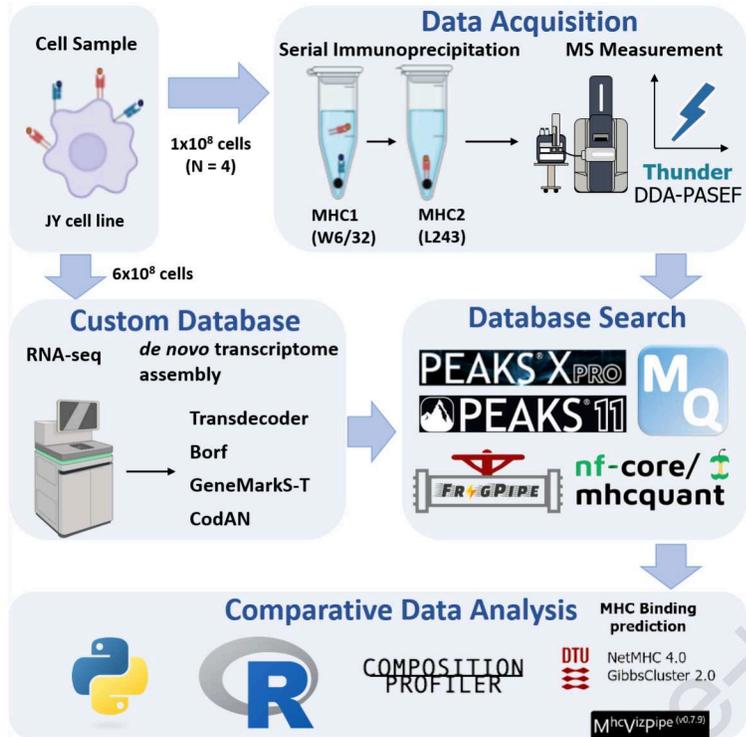
Revised Date: 1 December 2025

Accepted Date: 16 December 2025

Please cite this article as: Chen Y, Preikschat A, Arnold A, Pecori R, Gomez-Zepeda D, Tenzer S, Benchmarking Software for DDA-PASEF Immunopeptidomics, *Molecular & Cellular Proteomics* (2026), doi: <https://doi.org/10.1016/j.mcpro.2025.101492>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2025 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology.



Benchmarking Software for DDA-PASEF Immunopeptidomics

Authors:

Yannic Chen^{1,2}, Annica Preikschat³, Annette Arnold², Riccardo Pecori², David Gomez-Zepeda^{1,2},
Stefan Tenzer^{1,2,3,4}

Affiliation:

1. Helmholtz Institute for Translational Oncology Mainz (HI-TRON Mainz) – A Helmholtz Institute of the DKFZ, Mainz, Germany
2. DKFZ German Cancer Research Center, Heidelberg, Germany
3. University Medical Center Mainz, Mainz, Germany
4. Research Center for Immune Therapy [Forschungszentrum für Immuntherapie (FZI)], University Medical Center (UMC) of the Johannes Gutenberg University, Mainz, Germany

Correspondence: Stefan Tenzer (tenzer@uni-mainz.de)

Running title

Benchmarking Software for DDA-PASEF Immunopeptidomics

Abbreviations

DDA: Data Dependent Acquisition

DIA: Data Independent Acquisition

FDR: False Discovery Rate

MHC: Major Histocompatibility Complex

MHC1p: Major Histocompatibility Complex 1 peptide

MHC2p: Major Histocompatibility Complex 2 peptide

MS: Mass spectrometry

PASEF: Parallel Accumulation and Serial Fragmentation

PSM: Peptide-Spectrum Match

SAV: Single Amino acid Variants

SNV: Single Nucleotide Variants

TIMS: Trapped Ion Mobility Separation

Abstract

Mass spectrometry (MS) is the method of choice for high-throughput identification of immunopeptides, which are generated by intracellular proteases, unlike proteomics peptides that are typically derived from trypsin-digested proteins. Therefore, the searching space for immunopeptides is not limited by proteolytic specificity, requiring more sophisticated software algorithms to handle the increased complexity. Despite the widespread use of MS in immunopeptidomics, there is a lack of systematic evaluation of data processing software, making it challenging to identify the optimal solution. In this study, we provide a

comprehensive benchmarking of the most widespread/used data-dependent acquisition (DDA)-based software platforms for immunopeptidomics: MaxQuant, FragPipe, PEAKS and MHCquant. The evaluation was conducted using data obtained from the JY cell line using the Thunder-DDA-PASEF method. We assessed each software's ability to identify immunopeptides and compared their identification confidence. Additionally, we examined potential biases in the results and tested the impact of database size on immunopeptide identification efficiency. Our findings demonstrate that all software platforms successfully identify the most prominent subset of immunopeptides with 1% false discovery rate (FDR) control, achieving medium to high identification confidence correlations. The largest number of immunopeptides were identified using the commercial PEAKS software, which is closely followed by FragPipe, making it a viable non-commercial alternative. However, we observed that larger database sizes negatively impacted the performance of some software platforms more than others. These results provide valuable insights into the strengths and limitations of current MS data processing tools for immunopeptidomics, supporting the immunopeptidomics/MS community in determining the right choice of software.

Introduction

The Major Histocompatibility Complex (MHC) plays a crucial role in the adaptive immune response by presenting immunopeptides to T-cells, enabling the identification of aberrant cancerous and pathogenic non-self-peptides. MHC are classified into class 1 and class 2, with the former presenting primarily endogenous peptides derived from intracellular proteins while the latter primarily presents exogenous peptides acquired through endocytosis [1][2]. In

humans, this protein is termed human leukocyte antigen (HLA) and typically presents peptides of length 8-13 amino acids for class 1 and 13-25 for class 2 [3][4]. The study of the whole set of immunopeptides, known as the immunopeptidome, is fundamental to many medical research fields, particularly immunotherapy, where disease-specific immunopeptides can be leveraged for targeted treatment [1][5][6].

Mass spectrometry (MS) is the method of choice for high-throughput identification of immunopeptides[7]. MS techniques infer peptide sequences by comparing observed spectra to hypothetical peptide fragmentation pattern. However, this approach faces challenges due to missing fragmentation from incomplete fragmentations as well as interfering peaks from contaminants/other biological sources in case of multiplexed spectra. To address this, database search algorithms are employed to make informed predictions of peptide identities by limiting the possible identities to known peptide sequences [8]. It is important to note that peptides may undergo post-translational modifications (PTMs), which alter the fragment masses and expand the range of potential identities. To account for false positives inherent in this approach, False Discovery Rate (FDR) analysis is used as a quality control measure. FDR employs "decoy" databases containing sequences not present in the sample of interest, to assess the reliability of identifications by quantifying the rate of false "decoy hits" [9].

Immunopeptidomics poses unique challenges for database search engines compared to traditional trypsin-based bottom-up proteomics. In contrast to tryptic digests, immunopeptides are generated by intracellular proteases, necessitating searches without restrictions imposed by proteolytic specificities, dramatically expanding the search space and increasing computational demands [1].

Among various modes of measurement, Data-Dependent acquisition mode (DDA) is the preferred choice for high confidence identifications as it selects the most abundant ions by mass-to-charge ratio (m/z), thereby obtaining high-quality spectra [1][10].

Recently, tailored acquisition methods on the timsTOF instruments have resulted in high-coverage immunopeptidome profiling of biological samples. These instruments feature a dual TIMS (trapped ion mobility separation) setup, enabling parallel accumulation and serial fragmentation (PASEF)[11]. In this approach, peptides are accumulated in packages and then separated based on their ion mobility, significantly enhancing immunopeptide detection[12][13][14]. However, little attention has been given to evaluating search engines for LC-IMS-MS/MS (Liquid Chromatography Ion Mobility Mass Spectrometer) immunopeptidomics data acquired using DDA-PASEF, a hybrid methodology that integrates DDA with PASEF to achieve high-coverage data acquisition. In addition, many current programs were developed for general proteomics or tested on data from different instruments (e.g., Orbitrap).

To address this gap, we benchmarked database search algorithms for their performance in identifying immunopeptides from an optimized Thunder-DDA-PASEF acquisition strategy [14]. The search engines SEQUEST, MASCOT, and X!Tandem played a foundational role in the development of modern proteomics research. This study, however, focuses on the most current and relevant software available to date. MaxQuant [15], which uses the Andromeda search engine, is one of the most widely used tools for proteomics analysis. MSFragger from FragPipe [16] focuses on speeding up analysis by utilizing fragment ion indexing, making it ideal for searching large search spaces, such as those found in immunopeptidomics. Likewise, PEAKS [17] focuses on *de novo* assisted search, by generating tags for reducing the search space and

yielding more robust results. The newly released MHCquant [18], built around the Comet search engine, optimizes workflows for immunopeptidomics by refining result via downstream processes such as rescoring. Analysis of the degree of consensus among the different tools serves as an indicator of the individual result's reliability and quality. Finally, these insights will guide the selection of optimal software tools for specific research contexts.

Experimental Procedures

Experimental design and Statistical Rationale

To generate the benchmarking dataset, MHC class 1 (MHC1p) and MHC class 2 peptides (MHC2p) were enriched from the JY lymphoblastoid cell line using a sequential immunoprecipitation method (See below). Four technical replicates were analyzed via three injections each ($n = 12$) by LC-IMS-MS employing the Thunder-DDA-PASEF method. For MHC2p analysis, the isolation polygon was modified to accommodate multiply charged ions resulting from longer peptides in the multiply charged ion cloud.

The JY MHC1p and MHC2p datasets were processed using PEAKS X Pro, PEAKS 11, FragPipe (v.18.0 unless otherwise stated), MaxQuant (2.4.4.0), and MHCquant (2.6.0dev). Machine learning-based rescoring methods have been shown to enhance immunopeptide identification from DDA-PASEF data [14][19]. Since PEAKS 11 Studio and Online, and FragPipe include built-in rescoring, these options were activated for this analysis (deep-learning boost and MSBooster [20], respectively). Additional downstream rescoring third party tools for PEAKS X Pro and MaxQuant were not employed in order to compare software performance in their native state.

The heterogeneity in search engines, setting options, and rescoring algorithms across software packages complicates uniform comparison. To mitigate this, variables were standardized where possible (see below), with remaining parameters left at default. To address the quality of matching ambiguous spectra to peptide sequences, 1% PSM FDR filtering is employed, unless otherwise stated.

JY cell culture

The EBV (Epstein-Barr virus) transformed human B lymphoblastoid cell line JY (RRID:CVCL_0108)[21] was cultured in RPMI-1640 medium (Sigma-Aldrich, Cat# R8758) supplemented with 10% fetal bovine serum (PAN Biotech, Cat# P40-37100, v/v), 1 mM sodium pyruvate, 2 mM glutamine and 1% of Penicillin/Streptomycin (Sigma-Aldrich, Cat# P4333, v/v). Suspension cell cultures were harvested at 220 x g for 10min and washed three times with 1x PBS prior counting and freezing at -80°C .

Immunopeptide enrichment by immunoprecipitation (IP)

In brief, cell pellets (4×10^8 cells) were defrosted and lysed by adding cold 1% CHAPS (m/v) lysis buffer in PBS, shaking on ice, and sonicating the samples with a Sonicator (Bandelin Sonopuls). Pooled samples were re-distributed into four replicates (1×10^8 cells each) and immunoprecipitated overnight at 4°C rotating with anti-pan-HLA1 antibody (W6/32, Leinco H263) coupled to CNBr activated beads (Cytivia). A second immunoprecipitation of the unbound fraction was performed with anti-HLA-DR antibody (L243, Leinco H261) coupled beads instead. Beads were washed 2 times with 1xPBS and once with H_2O , followed by seven iterations of elution from beads with 200 μl 0.2% trifluoroacetic acid (TFA) shaking on ice.

Subsequently samples were ultrafiltrated in tubes with 10 kDa molecular weight cut off (MWCO) (Vivacon 500, Hydrosart) and desalted on Oasis HLB plate (Waters) using an elution buffer of 35% acetonitrile (ACN), 0.1% TFA. Finally, samples were lyophilized and dissolved in 15 μ l 0.1% formic acid (FA) shortly before LC-MS/MS analysis.

LC-MS/MS analysis

Desalted peptides were analyzed by nanoLC-MS using a nanoElute coupled to a timsTOF-Pro-2 mass spectrometer using Thunder-DDA-PASEF for MHC1p [14], with adaptations for MHC2p. Chromatographic separation was performed in a C18 reversed-phase column (Aurora 25 cm x 75 μ m ID, 120 \AA pore size, 1.7 μ m particle size, IonOpticks, Australia). The samples were injected directly into the analytical column and separated using a 47-minute gradient, increasing the proportion of phase B (ACN with 0.1% FA (v/v)) to phase A (water with 0.1% FA (v/v)). The gradient started at 2% B, increased to 17% within 23 min, then to 25% in 11.5 min, to 37% in 3.8 min, and to 95% in 3.8 min, and finished with a wash step of 4.9 min at 95% B. Ionization took place using a Captive Spray source, with a capillary voltage of 1600 V, dry gas at 3.0 L/min, dry temperature at 180 $^{\circ}$ C, and TIMS-in pressure of 2.7 mBar. The MS was piloted using Compass Hystar(v5.1) and timsControl (v.4.0.5) (Bruker). The m/z acquisition range was set at 100 to 2000 Th, and the high-sensitivity detection mode was activated. TIMS was set to a range of 0.65 to 1.75 Vs/cm² with accumulation and ramp time of 300 ms. The number of MS2 frames per cycle was set to three with an overlap of one. The fragmentation intensity threshold was set at 1000 and the target intensity at 20,000. For MHC1p, the Thunder isolation polygon was used to select precursors for fragmentation within the $1/K_0$ vs. m/z range of peptides with 8 to 13 amino acids, including singly and multiply charged ions [14]. For MHC2p, the isolation

polygon was modified to include all multiply charged ions and singly charged ions above 550 m/z.

RNA-sequencing

RNA was extracted from 6×10^6 JY cells using the AllPrep DNA/RNA Mini kit (QIAGEN, Cat# 80204) and treated with DNase (Invitrogen, Cat# AM 1907). RNA-seq libraries were prepared using Illumina's TruSeq Stranded mRNA Library Prep Kit. Libraries were sequenced with the Illumina NovaSeq 6K PE 100 S4 technology.

Custom databases

Unless otherwise stated, the standard reference database used is the UniProt human reference proteome database (retrieved from February 2020) extended by concatenating common contaminant sequences (total sequences = 20535).

For the *Arabidopsis* entrapment strategy, the final search database was generated by concatenating the above standard reference database with the UniProt *Arabidopsis thaliana* (UP000006548) reference database (retrieved in June 2023) (total sequences = 48033).

For database size comparison, an updated UniProt human reference proteome was used (retrieved in June 2023). Furthermore, additional protein sequences were included based on RNA-seq data. Total RNA sequencing data from JY cells were quality controlled by Adapter trimming using TrimGalore (0.6.10) [22] and duplicate removal using clumpify (BBMap 39.01) [23]. The reads are aligned using STAR (2.7.10b)[24] with default settings on hg38 assembly and then summarized using featureCounts (Subread 2.0.6) [25] based on UCSC annotation file

retrieved from June 2023. Transcripts (including isoforms) with any aligned reads are included in the database (total sequences = 74261).

De novo transcriptome assembly (DNTE) was done in Galaxy. Reads were adapter-trimmed using Trimmomatic (Galaxy Version 0.38.1) [26] and duplicates were not removed. rnaSPAdes (Galaxy Version 3.15.4) [27] on paired-end reads with default settings was used for DNTE.

Coding sequences were predicted using four softwares: transdecoder (Galaxy Version 5.5.0) [28], Borf (v1.3.0) [29], GeneMarkS-T [30] and CodAN (v1.2) [31], with default settings and translated. The same set of contaminants are always included (total sequences = 602631).

DDA database search

Raw MS data files from timsTOF (Bruker) mass spectrometer were loaded into the software as individual samples if possible (PEAKS X Pro and PEAKS 11 Online). Option to set different Bioreplicates (FragPipe) and Experiments (MaxQuant) were used and Match between runs was disabled. Unless otherwise stated, the same settings were used for both MHC1 and MHC2 samples. The Precursor Mass Tolerance and Fragment Mass Tolerance were set to 15ppm and 0.03 da respectively. The peptide length was set to 7-25 peptides. For PEAKS X Pro this was not possible and results were manually filtered for 7-25 length peptides. Peptide mass range and charges were set to 200-5000 and 1-4 respectively (FragPipe, MaxQuant and MHCquant). No fixed modification was set and variable modifications includes Oxidation (M), Acetylation (Protein N-term) and Cysteinylation, with a maximum of three variable modifications on peptides. Any build-in rescoring algorithm was kept on (MSbooster/percolator for FragPipe, Deep learning boost for PEAKS 11 Online and percolator/deeplc/ms2pip for MHCquant). Only

the Top PSM for each spectrum is sent for rescoring. Enzyme was set to none and digestion mode to unspecific/nonspecific. No extra contaminant Fasta file was provided (however, MaxQuant build-in contaminant was kept on as default). Regarding quantification, MHCquant only worked when the `--skip-quantification` option was enabled. For FragPipe, MS1 quant was enabled when quantification is needed.

Unless otherwise stated, the Protein and Peptide FDR were set to 1 or 100%, but the PSM FDR was left at 0.01 or 1% for all but PEAKS X Pro (which doesn't allow this setting), in which case the 1% PSM FDR was calculated manually (see statistics section). When retrieving all unfiltered candidate PSMs, PSM FDR was also set to 1 or 100%.

For FragPipe, when specifying different Bioreplicates, a file of the combined PSM result is not generated. This is generated by manually concatenating all `psm.tsv` files in each "exp" folder together. Unique peptide sequences are obtained by grouping peptide sequences together and keeping the entry with the highest "PeptideProphet.Probability" value. In the same way, a file containing all unique peptidofoms can be generated by using the "Modified Peptide" column instead of the "Peptide" column.

The identification scores used for PEAKS X Pro and PEAKS 11 Online is "X.10LgP" (higher is more confident), for FragPipe is "PeptideProphet.Probability" (higher is more confident), for MaxQuant "PEP" is used (lower is more confident) and MHCquant is "score" (lower is more confident).

Peptide binding prediction

Peptide binding prediction was done using netMHCpan version 4.1 [32] for MHC class 1 and netMHCIIpan version 4.1 [33] for MHC class 2 peptides. For JY cells, the MHC class 1 alleles HLA-A:02:01, HLA-B:07:02 and HLA-C:07:02 were used with a length between 8-13 amino acids. For MHC class 2 DRB1:04:04, DRB1:13:01, DRB3:01:01 and DRB4:01:03 were selected with a length of 9-22 amino acids. The rank threshold for determining strong and weak binders was kept at default.

Spectrum ID comparison

Both spectrum ID and source files are used as unique identifier (except for MHCquant). For PEAKS X Pro, PEAKS 11 Online and Fragpipe software, this information is provided in the psm data file. For Fragpipe specifically, the “Spectrum” column contains the source file and spectrum ID concatenated within a single string. For MaxQuant, the “accumulatedMsmsScans.txt” file is used, where the “PASEF precursor ID” is the equivalent to spectrum ID. MHCquant does not output spectrum ID in their results files. For comparison between itself, the RT and m/z values are used as unique identifier.

Entrapment strategy

Database search analysis was conducted as described in the DDA database search section, with the exception of employing the entrapment database and omitting FDR filtering. Subsequently, results were manually filtered to achieve a 1% peptidoform FDR. Matches exclusively associated to decoy accessions were removed. Target hits were defined as peptide hits corresponding to any protein accessions from the standard reference database, while

Arabidopsis false positives were classified as identified peptides associated exclusively to the Arabidopsis proteome. For comparative analysis, with the exception of MHCquant, target-decoy results were generated using an analogous approach, with decoy hits serving as false positives and excluding any matches exclusively linked to Arabidopsis protein accessions. MHCquant does not return decoy sequences and thus its 1% peptide FDR filtered results are used instead. Due to MHCquant including PTMs in its calculation its results are analogous to our definition of peptidoform.

Statistics

The upset plot of shared peptides between software was generated using ComplexUpset (version 1.3.3). Sankey graphs were generated using networkD3 (version 0.4.1).

Amino acid bias is analyzed with the online tool Composition Profiler [34]. The background Sample is the aggregate of the identified peptides from all software without removing duplicates. The individual results from each software is compared against this background sample. Significance threshold is set at 0.005 with other settings kept at default.

Spearman correlation coefficient was calculated by ranking the peptides based on the value used to determine their identification confidence (or peptide binding rank in case of binding prediction), unless otherwise stated. Rankings are ordered in a confidence descending order.

The pairwise plot is generated using GGally (version 2.2.1).

For software that did not provide built-in PSM FDR filtering capabilities, manual FDR calculation was performed. This process utilized the unfiltered PSM files generated by the software. FDR was calculated as follows:

$$FDR = \frac{Decoy_{CST}}{Total_ID_{CST}}$$

Where $Decoy_{CST}$ and $Total_ID_{CST}$ is the number of decoy and total identification remaining after filtering at confidence score threshold CST respectively. The 1% PSM FDR is determined by calculating the FDR at successive confidence score thresholds, moving from the most to least confident scores, and identifying the point where the $FDR \leq 0.01$.

The number of unique 9mers is found by splitting all sequences into all possible 9mers and filtering out duplicates.

Results

A note on Peaks Studio 11 and Peaks Online 11

Bioinformatics Solutions Inc. released PEAKS Studio 11 and PEAKS Online 11 in 2023, succeeding PEAKS X (Pro). Both use identical algorithms, with PEAKS Online 11 optimized for high-throughput processing via multiple nodes [35]. We confirmed that both produce identical results given the same inputs and settings. Only minor variations were observed when activating the deep learning boost function (Figure S1), attributable to the inherent randomness of neural network models across different hardware. Due to this redundancy, only PEAKS 11 Online results were included in further analysis.

The choice of search engine affects peptide identification from timsTOF Pro immunopeptidomics experiments

Software performance was primarily evaluated by the number of unique peptides identified at 1% PSM FDR. Here, we refer to peptides as distinct amino acid sequences and peptidoforms as PTM modified peptide variants.

Results (Figure 1A) show that four of the five software demonstrate comparable number of identified peptides and peptidoforms. PEAKS 11 Online has the highest number of identifications, with an increase of 6-11% for both MHC class 1 and 2 compared to PEAKS X Pro and FragPipe. MHCquant saw a reduction of 6-9% compared to PEAKS X Pro and FragPipe. MaxQuant identified the fewest peptides and peptidoforms, finding 60% (MHC class 1) to 50% (MHC class 2) less compared to MHCquant. The PTM distribution (Table S1) show overall similar trends across the software except for MHCquant where it has lower identification for cysteinylated MHC1 peptide and Acetylated MHC2 peptides, while also not being able to identify peptides with both acetylation and cysteinylation.

We also assessed the impact of stricter parameters (10 ppm and 0.02 Da and length restriction of HLA1: 8-13mer & max. 1700 Da, HLA2: 15-25mer & max. 2900 Da) on the identification numbers. We observed an overall reduction of identification when using the stricter setting. However, with a fairer comparison of only the number of peptides within the length restriction, we see that for HLA1, the number of 8-13mer identification is increased for PEAKS X Pro and MHCquant, while for HLA2, there is an increased number of identified 15-25mer for Peaks 11 Online, Fragpipe and MHCquant (Table S2).

To evaluate identification consistency across software, we compared shared peptide sequences (Figure 1B). 25% of MHC1p and 32% of MHC2p were identified by all software. Excluding MaxQuant, this increases to a total of 58 and 60%, respectively. MaxQuant results show the lowest absolute overlap with other software, primarily due to its lower number of identifications rather than identifying different peptides. Only between 2-10% of identified peptides are software specific, indicating general consensus in identifications. To evaluate whether the observed peptide sequence identification consensus extends to the PSM level, we examined whether identical spectrum IDs were assigned to the same peptide sequences. We observed a substantial overlap in spectrum IDs linked to the same sequences between the two PEAKS software versions and MaxQuant. In contrast, most PSMs reported by FragPipe were associated with different spectrum IDs, and more than half of the spectrum IDs it shared with other tools were assigned to different peptide sequences (Figure S2).

The precursor charge and peptide length distribution is largely conserved between search engines

Beyond sensitivity, assessing peptide identification biases is crucial for validating software performance. To this end, we compared the precursor charge and peptide length distributions across search engines. Charge distributions were largely consistent, with most identifications carrying a +2 charge. As expected, MHC class I peptides contained a high fraction of singly charged precursors with relatively few triply charged peptides, whereas MHC class II peptides displayed the opposite trend (Figure 1C). FragPipe and MHCquant did not identify the very rare +4 charged peptides. These results indicate that all tools reproduce the characteristic charge state patterns of MHC-bound peptides, suggesting that no systematic bias in charge assignment

is introduced by the search engines. Our results show that singly charged immunopeptides contribute significantly to the total number of identifications (Figure S3). Notably, 76% - 85% of singly charged MHC1 peptides identified in our study were not identified with higher charge states, for MHC2 this proportion is even higher (85% - 90%). This indicates that most singly charged peptides would not be identified by acquisition methods targeting only multiply charged ions, consistent with the observations of Gomez-Zepeda *et al.* (2024).

In immunopeptidomics, the length distribution of identified peptides serves as a key indicator of immunopeptide enrichment but also identification performance. Indeed, MHC1p typically have a length of 8 to 13 amino acids with a large predominance of 9-mers[3], while MHC2p extend up to 25 amino acids with a length distribution peaking at 15-mers (in humans)[36]. The length distribution is largely consistent across software packages (Figure 1D). The enrichment of 9-mers in the MHC class 1 and length distribution of MHC class 2 samples aligns with expectations and suggests successful immunopeptide enrichment and identification. The presence of 9-mers in the MHC class 2 sample may indicate a co-enrichment of MHC class 1 peptide complexes during immunoprecipitation [37]. However, MaxQuant demonstrates reduced efficiency in identifying peptides shorter than 9 amino acids. For MHC class 2, this bias is further evidenced by increased identification of longer peptides.

MHC Binding Prediction indicates a Yield vs. Specificity Trade-off

Complementing on our analysis of peptide length distributions, we further evaluated software accuracy by leveraging predicted MHC binding affinity. This approach allows for a more

comprehensive assessment of software performance in identifying biologically relevant peptides.

We used netMHCpan [32] and netMHCIIpan[33] to predict binding affinities for identified peptides (Figure 2A). The HLA alleles used are described in the Methods section. For MHC class 1 peptides, PEAKS X Pro and 11 showed similar fractions of predicted binders (0.550 and 0.548 respectively). The other three software packages demonstrated slightly higher fractions, with MHCquant performing best (0.621). For MHC class 2, MaxQuant showed the highest fraction of predicted binder (0.742). The other four software packages performed comparably. PEAKS 11 has slightly lower proportion of predicted binders, but it has the highest total number of predicted binders (10261 MHC1p and 16012 MHC2p), followed by FragPipe (9674 MHC1p and 15750 MHC2p). The increased identification of non-binders by certain software could point to greater sensitivity, since such peptides are expected to be less abundant after purification. However, binding predictions are not fully accurate, and some of these non-binders may in fact represent true ligands that are not recognized by the prediction model or that bind to non-classical HLA molecules (e.g., HLA-E, HLA-G). Alternatively, they could also arise from contaminants or misassigned spectra.

To investigate whether strong MHC binders are more consistently identified, we examined the relationship between predicted binding affinity and the number of search engines reporting each peptide. As shown in Figure 2B, peptides with stronger predicted binding affinities were consistently identified by multiple search engines for both MHC class I and class II across all alleles. Nevertheless, Figure S4 shows that non-, weak-, and strong binders are all frequently

identified, suggesting that predicted non-binders (perhaps non-MHC peptides) may indeed be present in the measured sample.

Amino acid composition is largely conserved between search engines

Parker *et al.* (2021) previously reported that search engines may prioritize different peptides based on their amino acid composition, which would introduce a bias in immunopeptide identification [38]. To evaluate if this is the case with the software tested here for DDA-PASEF data, we employed Composition Profiler [34] to quantify significant variations. This software tests whether certain amino acids or properties are enriched or depleted in a dataset compared to a background population. Here, we used the aggregate of all identified peptides as background. This approach provides a robust baseline, capturing the full spectrum of identifications and reflecting software-specific biases in amino acid composition of identified peptides, independently of other factors such as binding-motif specific enrichment. It's important to note that this method introduces another bias, as software identifying more peptides contributes more to the background population.

Analysis of the amino acid distribution of identified peptides revealed largely conserved patterns across search engines. Notably, MaxQuant displayed the greatest number of significant deviations in amino acid frequencies for both MHC class I and II peptides (Figure 2C, Figure S5). Consistently, MaxQuant also showed the most frequent significant differences in broader amino acid properties, such as charge and polarity (Table S3). These findings suggest that MaxQuant's peptide identification algorithm introduced the most distinct biases in the amino acid composition of reported peptides compared to other software packages. It must be

noted that for MHC class 1, MHCquant has nearly the same number of significantly enriched/depleted properties as MaxQuant. Finally, visualization of motifs using unsupervised Gibbs Clustering by GibbsCluster [39] showed no obvious differences. (Figure S6). Altogether, these results indicate that, despite the slight differences, all the programs identified peptides with similar MHC binding motifs.

Missed consensus Peptides

To further explore differences between search engines, we defined “consensus peptides” as those identified by at least three database search engines. We then examined the properties of consensus peptides that were not detected by each individual software (Figure S7).

Interestingly, a large fraction of the consensus peptides missed by FragPipe for both MHC class I and II are at the shorter end of the length distribution. This appears to be advantageous for MHC class II, as many of these missed consensus peptides are predicted non-binders.

Reproducibility and Quantification

To compare the reproducibility of the search engines across the measured samples, we generated a data completeness curve (Figure 2D). The result show similar data completeness with MHCquant performing slightly better while MaxQuant slightly worse for MHC1 peptides.

We further assessed peptide quantification reproducibility by calculating the coefficient of variation (CV) for each peptide (area for PEAKS X Pro and PEAKS 11 Online and intensity for FragPipe and MaxQuant. MHCquant quantification was skipped due to software issues) across samples and found that MaxQuant exhibited the largest CV distribution, indicating higher variability in peptide quantification (Figure S8A).

Finally, we compared peptide quantification between search engines to assess consistency in relative abundance. Figure S8B shows strong correlations across all software, based on Spearman rank correlation coefficients, demonstrating a high degree of agreement for peptides identified by multiple tools.

Peptide identification scores correlate across software

Comparing the scores of identified peptides between software is crucial for revealing discrepancies in confidence levels and ranking of matches, potentially uncovering algorithmic biases that may be hidden when only considering overlap. However, direct comparison of absolute scores is challenging due to varying scoring methods and the impact of rescoring/boosting, which often results in bimodal score distributions. Therefore, to assess result similarity, we calculated pairwise Spearman correlation coefficients (Figure 3A). Results show moderate to strong correlations between software. PEAKS X pro and FragPipe show the most similar confidence levels with a spearman correlation score of 0.8, while the lowest correlation can be observed between MaxQuant and Peaks 11 Online. Using the relative rankings, we observed that peptide with better ranks are consistently found across multiple software (Figure 3B).

Following the observations in Figure 2B, we hypothesized that higher-affinity peptides could more likely persist through the sample preparation and accumulate for better MS identification. To this end, we evaluated the correlation between the Rnk_EL score and each software score rank. This however showed little to no correlation (Figure S9).

These findings underscore the complex interplay between peptide identification algorithms, scoring methods, and biochemical factors in immunopeptidomics, highlighting the need for cautious interpretation of software confidence scores and their relationship to immunopeptide identification.

To ensure clarity, we focus the remaining sections on MHC class 1 peptides. As of writing, MHC 1 reflects the current interest trend in the immunopeptide research community, allowing for a more manageable scope and a deeper, more consistent evaluation of search engine performance.

Arabidopsis proteome entrapment sequence indicate no bias in FDR control strategy. FDR filtering using decoys helps mitigate false positives. However, unbiased implementation of this can be difficult. An effective way to assess the robustness of each software's FDR implementation involves implementing an entrapment strategy with externally supplied entrapment sequences that act as an independent set of false positives to the software generated decoy sequences. This method helps detect potential biases from the software own FDR control implementations and provides a more comprehensive and independent evaluation of software results. Here we used *Arabidopsis thaliana* proteome sequences as entrapment sequence.

In order to compare the performance between the *Arabidopsis* entrapment strategy and the native target-decoy strategy, a 1% peptidiform FDR cutoff was used for both approaches. We found that a 1% PSM FDR approximates to 1.5-3.2% peptidiform FDR (and ~5% peptide FDR if

quantifiable) depending on software (Table S4). This shows that peptidoform level FDR acts as a stricter quality control feature compared to PSM FDR.

The entrapment strategy yielded higher number of identifications at 1% peptidoform FDR for all software compared to their native target-decoy FDR control. For the software that could be tested, this gap in identification expanded as the FDR increased (Figure 4). This indicates that decoy hits are not inherently scored worse over the *Arabidopsis* false positive hits, suggesting that the native target-decoy implementations are unbiased and not overestimating FDR filtered identification.

Database size affect peptide identification

As we have seen in the previous section, the reference database is a critical parameter in proteomics searches, affecting consistency and accuracy of peptide identifications by influencing FDR filtering and scoring. Here we further investigate this by systematically testing different database size to reveal how software adapts to changing search spaces, exposing strengths and limitations not evident with a single fixed database.

To simulate real-world research scenarios, we crafted a series of reference databases of varying sizes, based on JY cell line RNA-seq data (see Methods and Figure 5B table). This approach serves a dual purpose: it minimizes the inclusion of irrelevant sequences likely absent in our samples, while simultaneously incorporating non-canonical isoforms and accounting for edits, mutations, and variants. The latter exemplifies common use-cases in immunopeptidomics, particularly when identifying novel targets of interest.

As database size increases, most software packages reported increase number of candidate PSMs (at 100% PSM FDR filtering), except for PEAKS, which surprisingly exhibits a decrease (Figure 5). This unexpected trend persists despite larger databases containing a superset of sequences from smaller ones, which should theoretically allow for more comprehensive peptide-spectrum matches. The cause remains undetermined despite consultation with Bioinformatics Solutions Inc. With 1% PSM FDR filtering, the results obtained with the largest database (UniProt + transcriptome + *de novo*) saw the largest reduction, to the point where the number of target identifications is the lowest for all software. Furthermore, we observed that the FDR cutoff value generally becomes stricter with larger databases (Table S5). However, this is not a perfect correlation, indicating that other aspects play a role here.

When expanding the databases, especially when including sequences derived from *de novo* transcriptome assembly (DNTA) into the database, we observed a high proportion of candidates PSMs derived from those in FragPipe, MHCquant and MaxQuant (Figure 5A lighter shaded area). However, most of these PSMs are excluded upon FDR filtering (Figure 5C).

The UniProt reference database consistently yielded the highest number of FDR filtered peptidofoms identifications for the PEAKS software, while the inclusion of the transcriptome resulted in most identifications for the other three software. The addition of sequences from the DNTA data resulted in the lowest number of identifications. This indicate that addition of sequences of interest requires careful consideration to avoid overly expanding the search space, which may negatively affect the results.

DNA/RNA sequencing enables detection of peptide variants absent from reference databases. Single nucleotide variants (SNVs), the most prevalent class of genetic variants, involves a single

nucleotide change that may correspond to a single amino acid substitution resulting in a single amino acid variant (SAV). We evaluated the ability of the software to distinguish SAVs through the identification of pairs of SAV and its corresponding reference. All tools successfully detected such pairs, with both PEAKS software reporting both the highest number as well as fractions of this type of variant, especially when the DNTA were included (Figure 5D both lighter shaded areas). However, many identified were leucine/isoleucine substitutions (Figure 5D lightest shaded area) —indistinguishable by mass—suggesting that a single spectrum may match multiple peptide candidates, depending on how results are reported. This observation raises questions about whether such cases constitute distinct identifications. In PEAKS 11 Online, 8 out of 13 and 32 out of 44 peptides are “IL variants” (for +transcriptome and + DNTA respectively). PEAKS X Pro follows a similar trend with 5/10 and 20/29 peptides. MHCquant has 4/10 and 8/20, MaxQuant has 2/3 and 2/2 and FragPipe has 0/5 and 3/16 IL variants. From this, PEAKS software should be the method of choice when “I/L variant” detection is relevant.

We were further interested in the non-reference identifications and examined their general properties (Figure 6). Compared to the findings displayed in Figure 1 and 2, the results here show much lower consistency in shared peptides, length and predicted binding distribution and binding prediction. A notable difference in the binding motif is that the characteristic acidic amino acid signal at position 4, a prominent feature of HLA-A:02:01, was largely absent.

To further investigate how PSMs are affected by database size, we compared MS1 spectrum IDs between the UniProt-only database and the expanded UniProt + transcriptome + DNTA database (Figure S10). For both PEAKS software, many PSMs shifted to a “worse” category: PSMs that previously passed the FDR filter no longer did, and PSMs that did not pass the filter

were no longer matched at all. Few new spectra were matched to peptides when using the UniProt + Transcriptome + DNTA database. In contrast, FragPipe and MHCquant matched many new spectra, explaining the gain in peptide identifications. Interestingly, MHCquant exhibited the largest changes among unique PSMs (those identified only by one of the search results) with most of these unique PSMs passing the FDR threshold. MaxQuant, unlike the other software, did not lose any spectrum IDs when switching to the larger database. The majority of spectra retained the same peptide assignment with high average scores. The spectrum that were either assigned to new sequences or left unassigned in one of the two search results (unique PSMs) generally had worse confidence scores, suggesting that high-confidence PSMs remain largely stable despite database expansion.

Discussion

We compared five software packages based on four algorithms for immunopeptide identification from the optimized Thunder DDA-PASEF LC-MS data acquired in a timsTOF Pro2 mass spectrometer. We demonstrate that all evaluated software packages can identify the most evident subset of immunopeptides. PEAKS 11 with the build-in rescoring algorithm exhibited superior performance in immunopeptidome identification under standard operating conditions. FragPipe emerged as a viable non-commercial alternative, yielding comparable identification results.

Software selection depends on specific research needs. Although PEAKS 11 Online identified the highest number of peptides, its performance declined with increasing database size, potentially limiting its efficacy for analyses involving extensive reference databases.

Correspondence with the PEAKS software staff revealed that this performance degradation may stem from the presence of similar sequences in larger databases. They caution against using reference databases containing numerous similar sequences, such as those incorporating novel SNPs and isoforms, as this may compromise their software's effectiveness¹. In such cases, FragPipe is a good choice. On the other hand, PEAKS return the most I/L variant peptides, which may be important for studies emphasizing such targets.

There is ongoing discussion that applying false discovery rate (FDR) control on a per-group basis (grouped FDR control) can increase identifications within specific groups [40]. In our view, this approach is most appropriate when there is a clear rationale for holding different subsets of identifications to separate statistical standards. For example, in multi-round searches where later rounds inherently operate on different quality spectra. This was not the case in our study, as all databases (both the entrapment approach and the database size comparison) were concatenated and searched in a single round.

The PEAKS 11th generation software offers an immunopeptide-specific workflow that employs a machine learning model tailored to immunopeptides. This particular approach was not incorporated in our analysis. Although this is a limitation of our study, here we focused on evaluating database search engines. Future research may benefit from a thorough examination of this workflow's performance and potential applications in immunopeptide analysis.

Machine learning-based rescoring functions have been shown to improve peptide identification under FDR filtering [41] in MaxQuant [19] and Peaks X Pro [14], which is also seen from our

¹ Kristina Jurcic (Bioinformatics Solutions Inc.), oral communication, 26th April 2024.

results comparing Peaks X Pro and Peaks 11 Online. The entrapment strategy demonstrated unbiased target-decoy performance across all software, including those utilizing deep learning algorithms. This finding is significant, addressing concerns about potential overfitting in deep learning-based methods. However, caution is warranted when using deep learning approaches due to inherent randomness, which may affect reproducibility. Thorough evaluation of deep learning models remains essential to identify potential biases that may impact result integrity. Our findings contribute to the optimization of immunopeptidome analysis workflows and aid in software selection for future studies in this field. Especially low quantity targets can prove challenging for identification, making the correct choice of bioinformatics analysis essential. In the future we look towards data-independent acquisition (DIA) strategies, as search engines mature and AI driven technology are incorporated. DIA allows for the simultaneous acquisition of all precursor ions within a specified mass range, improving the depth and comprehensiveness of proteomic analysis, but at the same time generates more complex spectra that require much more sophisticated processing and bioinformatic analysis.

Database search software versions

Peaks X Pro: 10.6 build 20201221

Peaks Online 11: September 2023 version

Peaks Studio 11: build 20230202

FragPipe: 18.0/ 3.5/ 4.4.0/1.8.0 (FragPipe/MSFragger/Philosopher/IonQuant) – FragPipe 21.2-build03 was used for database size comparison.

MaxQuant: 2.4.4.0

MHCquant: 2.6.0dev

CRedit authorship contribution statement

YC: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft, Visualization, Software. AP, AA and RP: Investigation. DGZ Conceptualization, Supervision, Writing – Review & Editing. ST: Funding acquisition, Supervision, Writing – Review & Editing.

Acknowledgements

We acknowledge the support of the Immunopeptidomics Platform of HI-TRON Mainz.

We would like to acknowledge Lucas Kleinort (HI-TRON, Mainz, Germany) for his technical assistance. We also thank the Sequencing Core Facility at the German Cancer Research Center (DKFZ) for providing excellent sequencing services.

AI Statement: During the preparation of this work, the authors utilized ChatGPT version 4 to enhance readability and conciseness. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Data availability

The mass spectrometry immunopeptidomics and proteomics data have been deposited to the ProteomeXchange Consortium [42] via the jPOSTrepo partner repository [43] with the dataset identifiers PXD065501 for ProteomeXchange and JPST003768 for jPOSTrepo.

Transcriptomic Data

The RNA-seq data for this study have been deposited in the European Nucleotide Archive (ENA) at EMBL-EBI under accession number PRJEB98909 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB98909>).

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) SFB1292 TP13 (H.S.) and TPQ01 (S.T.), the Federal Ministry of Education and Research (BMBF) (MSCoreSys, DIASyM, Fkz: 161L0218, 16LW0241K to S.T.), the HI-TRON Kick-Start Seed Funding Program 2021 awarded to R.P. and S.T., the Helmholtz-Institute for Translational Oncology Mainz (HI-TRON Mainz) - a Helmholtz institute by DKFZ, Mainz, Germany.

Supplemental data

This article contains supplemental data.

Competing interests

Conflicts of interest: none

References

- [1] B. Zhang and M. Bassani-Sternberg, 'Current perspectives on mass spectrometry-based immunopeptidomics: The computational angle to tumor antigen discovery', Oct. 29, 2023, *BMJ Publishing Group*. doi: 10.1136/jitc-2023-007073.
- [2] K. L. Rock, E. Reits, and J. Neefjes, 'Present Yourself! By MHC Class I and MHC Class II Molecules', Nov. 01, 2016, *Elsevier Ltd*. doi: 10.1016/j.it.2016.08.010.
- [3] T. Trolle *et al.*, 'The Length Distribution of Class I-Restricted T Cell Epitopes Is Determined by Both Peptide Supply and MHC Allele-Specific Binding Preference', *J Immunol*, vol. 196, no. 4, pp. 1480–1487, Feb. 2016, doi: 10.4049/JIMMUNOL.1501721.
- [4] M. Wieczorek *et al.*, 'Major histocompatibility complex (MHC) class I and MHC class II proteins: Conformational plasticity in antigen presentation', Mar. 17, 2017, *Frontiers Research Foundation*. doi: 10.3389/fimmu.2017.00292.
- [5] C. Chong, G. Coukos, and M. Bassani-Sternberg, 'Identification of tumor antigens with immunopeptidomics', Feb. 01, 2022, *Nature Research*. doi: 10.1038/s41587-021-01038-8.
- [6] H. B. Taylor *et al.*, 'MS-based HLA-II peptidomics combined with multiomics will aid the development of future immunotherapies', Jun. 01, 2021, *American Society for Biochemistry and Molecular Biology Inc*. doi: 10.1016/J.MCPRO.2021.100116.

- [7] I. E. Shapiro and M. Bassani-Sternberg, 'The impact of immunopeptidomics: From basic research to clinical implementation', Mar. 01, 2023, *Academic Press*. doi: 10.1016/j.smim.2023.101727.
- [8] V. Marquioni, F. M. F. Nunes, and M. T. M. Novo-Mansur, 'Protein Identification by Database Searching of Mass Spectrometry Data in the Teaching of Proteomics', *J Chem Educ*, vol. 98, no. 3, pp. 812–823, Mar. 2021, doi: 10.1021/acs.jchemed.0c00853.
- [9] A. I. Nesvizhskii, 'Proteogenomics: Concepts, applications and computational strategies', Nov. 01, 2014, *Nature Publishing Group*. doi: 10.1038/NMETH.3144.
- [10] H. S. Pak *et al.*, 'Sensitive immunopeptidomics by leveraging available large-scale multi-HLA spectral libraries, data-independent acquisition, and MS/MS prediction', *Molecular and Cellular Proteomics*, vol. 20, 2021, doi: 10.1016/J.MCPRO.2021.100080.
- [11] F. Meier *et al.*, 'Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer', *Molecular and Cellular Proteomics*, vol. 17, no. 12, pp. 2534–2545, Dec. 2018, doi: 10.1074/mcp.TIR118.000900.
- [12] N. Hoenisch Gravel *et al.*, 'TOFIMS mass spectrometry-based immunopeptidomics refines tumor antigen identification', *Nat Commun*, vol. 14, no. 1, Dec. 2023, doi: 10.1038/s41467-023-42692-7.
- [13] K. M. Phulphagar *et al.*, 'Sensitive, High-Throughput HLA-I and HLA-II Immunopeptidomics Using Parallel Accumulation-Serial Fragmentation Mass Spectrometry', *Molecular and Cellular Proteomics*, vol. 22, no. 6, Jun. 2023, doi: 10.1016/j.mcpro.2023.100563.

- [14] D. Gomez-Zepeda *et al.*, 'Thunder-DDA-PASEF enables high-coverage immunopeptidomics and is boosted by MS2Rescore with MS2PIP timsTOF fragmentation prediction model', *Nat Commun*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-46380-y.
- [15] J. Cox and M. Mann, 'MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification', *Nat Biotechnol*, vol. 26, no. 12, pp. 1367–1372, Dec. 2008, doi: 10.1038/nbt.1511.
- [16] A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, and A. I. Nesvizhskii, 'MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics', *Nat Methods*, vol. 14, no. 5, pp. 513–520, Apr. 2017, doi: 10.1038/nmeth.4256.
- [17] L. Xin *et al.*, 'A streamlined platform for analyzing tera-scale DDA and DIA mass spectrometry data enables highly sensitive immunopeptidomics', *Nat Commun*, vol. 13, no. 1, Dec. 2022, doi: 10.1038/s41467-022-30867-7.
- [18] L. Bichmann *et al.*, 'MHCquant: Automated and Reproducible Data Analysis for Immunopeptidomics', *J Proteome Res*, vol. 18, no. 11, pp. 3876–3884, Nov. 2019, doi: 10.1021/acs.jproteome.9b00313.
- [19] C. Adams *et al.*, 'Fragment ion intensity prediction improves the identification rate of non-tryptic peptides in timsTOF', *Nat Commun*, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-48322-0.

- [20] K. L. Yang *et al.*, 'MSBooster: improving peptide identification rates using deep learning-based features', *Nature Communications* 2023 14:1, vol. 14, no. 1, pp. 1–14, Jul. 2023, doi: 10.1038/s41467-023-40129-9.
- [21] 'Cellosaurus cell line JY (CVCL_0108)'. Accessed: May 06, 2025. [Online]. Available: https://www.cellosaurus.org/CVCL_0108
- [22] F. Krueger, 'FelixKrueger/TrimGalore: A wrapper around Cutadapt and FastQC to consistently apply adapter and quality trimming to FastQ files, with extra functionality for RRBS data'. Accessed: Apr. 14, 2025. [Online]. Available: <https://github.com/FelixKrueger/TrimGalore>
- [23] B. Bushnell, 'BBMap | SourceForge.net'. Accessed: Apr. 14, 2025. [Online]. Available: <https://sourceforge.net/projects/bbmap/>
- [24] A. Dobin *et al.*, 'STAR: ultrafast universal RNA-seq aligner', *Bioinformatics*, vol. 29, no. 1, p. 15, Jan. 2012, doi: 10.1093/BIOINFORMATICS/BTS635.
- [25] Y. Liao, G. K. Smyth, and W. Shi, 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, vol. 30, no. 7, pp. 923–930, Apr. 2014, doi: 10.1093/BIOINFORMATICS/BTT656.
- [26] A. M. Bolger, M. Lohse, and B. Usadel, 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, vol. 30, no. 15, pp. 2114–2120, Aug. 2014, doi: 10.1093/BIOINFORMATICS/BTU170.

- [27] E. Bushmanova, D. Antipov, A. Lapidus, and A. D. Prjibelski, 'rnaSPAdes: a *de novo* transcriptome assembler and its application to RNA-Seq data', *Gigascience*, vol. 8, no. 9, pp. 1–13, Sep. 2019, doi: 10.1093/GIGASCIENCE/GIZ100.
- [28] B. J. Haas, 'TransDecoder'. Accessed: Apr. 14, 2025. [Online]. Available: <https://github.com/TransDecoder/TransDecoder/wiki>
- [29] B. Signal and T. Kahlke, 'Borf: Improved ORF prediction in de-novo assembled transcriptome annotation', *bioRxiv*, p. 2021.04.12.439551, Apr. 2021, doi: 10.1101/2021.04.12.439551.
- [30] S. Tang, A. Lomsadze, and M. Borodovsky, 'Identification of protein coding regions in RNA transcripts', *Nucleic Acids Res*, vol. 43, no. 12, pp. e78–e78, Jul. 2015, doi: 10.1093/NAR/GKV227.
- [31] P. G. Nachtigall, A. Y. Kashiwabara, and A. M. Durham, 'CodAn: predictive models for precise identification of coding regions in eukaryotic transcripts', *Brief Bioinform*, vol. 22, no. 3, pp. 1–11, May 2021, doi: 10.1093/BIB/BBAA045.
- [32] B. Reynisson, B. Alvarez, S. Paul, B. Peters, and M. Nielsen, 'NetMHCpan-4.1 and NetMHCIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data', *Nucleic Acids Res*, vol. 48, no. W1, pp. W449–W454, 2021, doi: 10.1093/NAR/GKAA379.
- [33] J. B. Nilsson *et al.*, 'Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning', 2023. [Online]. Available: <https://www.science.org>

- [34] V. Vacic, V. N. Uversky, A. K. Dunker, and S. Lonardi, 'Composition Profiler: a tool for discovery and visualization of amino acid composition differences', *BMC Bioinformatics*, vol. 8, Jun. 2007, doi: 10.1186/1471-2105-8-211.
- [35] 'Bioinformatics Solutions Inc. | LC-MS/MS Proteomics & Antibody Protein Characterisation'. Accessed: Aug. 09, 2024. [Online]. Available: <https://www.bioinfor.com//>
- [36] D. M. Tadros, S. Eggenschwiler, J. Racle, and D. Gfeller, 'The MHC Motif Atlas: a database of MHC binding specificities and ligands', *Nucleic Acids Res*, vol. 51, no. D1, pp. D428–D437, Jan. 2023, doi: 10.1093/NAR/GKAC965.
- [37] T. Partridge *et al.*, 'Discrimination between human leukocyte antigen class I-Bound and co-purified HIV-derived peptides in immunopeptidomics workflows', *Front Immunol*, vol. 9, no. APR, Apr. 2018, doi: 10.3389/fimmu.2018.00912.
- [38] R. Parker *et al.*, 'The choice of search engine affects sequencing depth and HLA Class I allele-specific peptide repertoires', *Molecular and Cellular Proteomics*, vol. 20, 2021, doi: 10.1016/J.MCPRO.2021.100124.
- [39] M. Andreatta, B. Alvarez, and M. Nielsen, 'GibbsCluster: unsupervised clustering and alignment of peptide sequences', *Nucleic Acids Res*, vol. 45, no. Web Server issue, p. W458, Jul. 2017, doi: 10.1093/NAR/GKX248.
- [40] J. Freestone, T. Short, W. S. Noble, and U. Keich, 'Group-walk: a rigorous approach to group-wise false discovery rate analysis by target-decoy competition', *Bioinformatics*,

vol. 38, no. Supplement_2, pp. ii82–ii88, Sep. 2022, doi:

10.1093/BIOINFORMATICS/BTAC471.

- [41] M. Kalhor, J. Lapin, M. Picciani, and M. Wilhelm, 'Rescoring Peptide Spectrum Matches: Boosting Proteomics Performance by Integrating Peptide Property Predictors Into Peptide Identification', *Molecular & Cellular Proteomics*, vol. 23, no. 7, p. 100798, Jul. 2024, doi: 10.1016/j.mcpro.2024.100798.
- [42] J. A. Vizcaíno *et al.*, 'ProteomeXchange provides globally co-ordinated proteomics data submission and dissemination', *Nat Biotechnol*, vol. 32, no. 3, p. 223, 2014, doi: 10.1038/NBT.2839.
- [43] S. Okuda *et al.*, 'jPOSTrepo: an international standard data repository for proteomes', *Nucleic Acids Res*, vol. 45, no. D1, pp. D1107–D1111, Jan. 2017, doi: 10.1093/NAR/GKW1080.

Figure legends

Figure 1. Comparative analysis of database search engine results. MHC class 1 results are in the left column and MHC class 2 results are in the right column. A) Bar chart showing the number of unique peptides identified by each software. Peptidoforms are included in lighter blue. B) Upset plot showing the intersection size of shared peptide sequences, ordered by intersection size. Only groupings resulting in an intersection size larger than 1% are included. C) Bar chart summarizing the charge distribution of identified peptides for each software. D) Column chart showing the relative length distribution of identified peptides for each software.

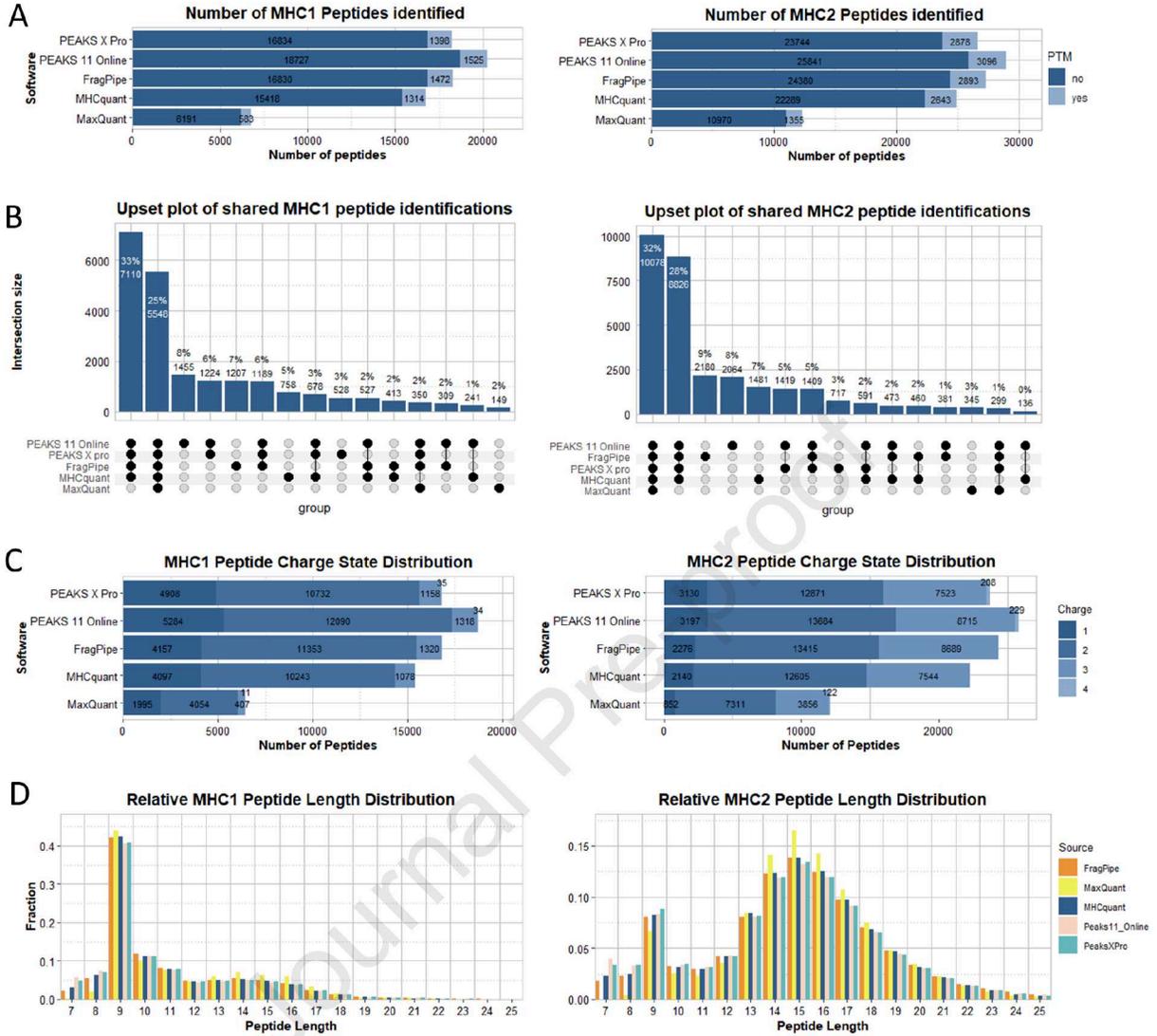
Figure 2. Analysis on peptide binding strength, amino acid composition and identification consistency. A) Bar chart showing the result of netMHCpan (left) and netMHCIpan (right) binding prediction results for the result of each software, showing the fraction of strong binders ($Rnk_EL \leq 0.05\%$ MHC class 1 and $\leq 1\%$ class 2), weak binders ($Rnk_EL > 0.05\%$ and $\leq 2\%$ for MHC class 1 and $> 1\%$ and $\leq 5\%$ for class 2) and non-binders ($Rnk_EL > 2\%$ for MHC class 1 and $> 5\%$ for class 2). B) Box and whisker plots comparing the rank score distribution of netMHCpan and netMHCIpan with the number of software the peptide was identified with. Blue dashed lines indicate the threshold for weak binders and the red dashed line indicates threshold for strong binder. C) Heatmap of the amino acid composition of the peptides identified by the software. Grey indicates not significant based on CompositionProfiler, with a significance threshold of 0.005. Values indicate the percentage difference from the background. D) Data completeness plot for each software across the 12 samples.

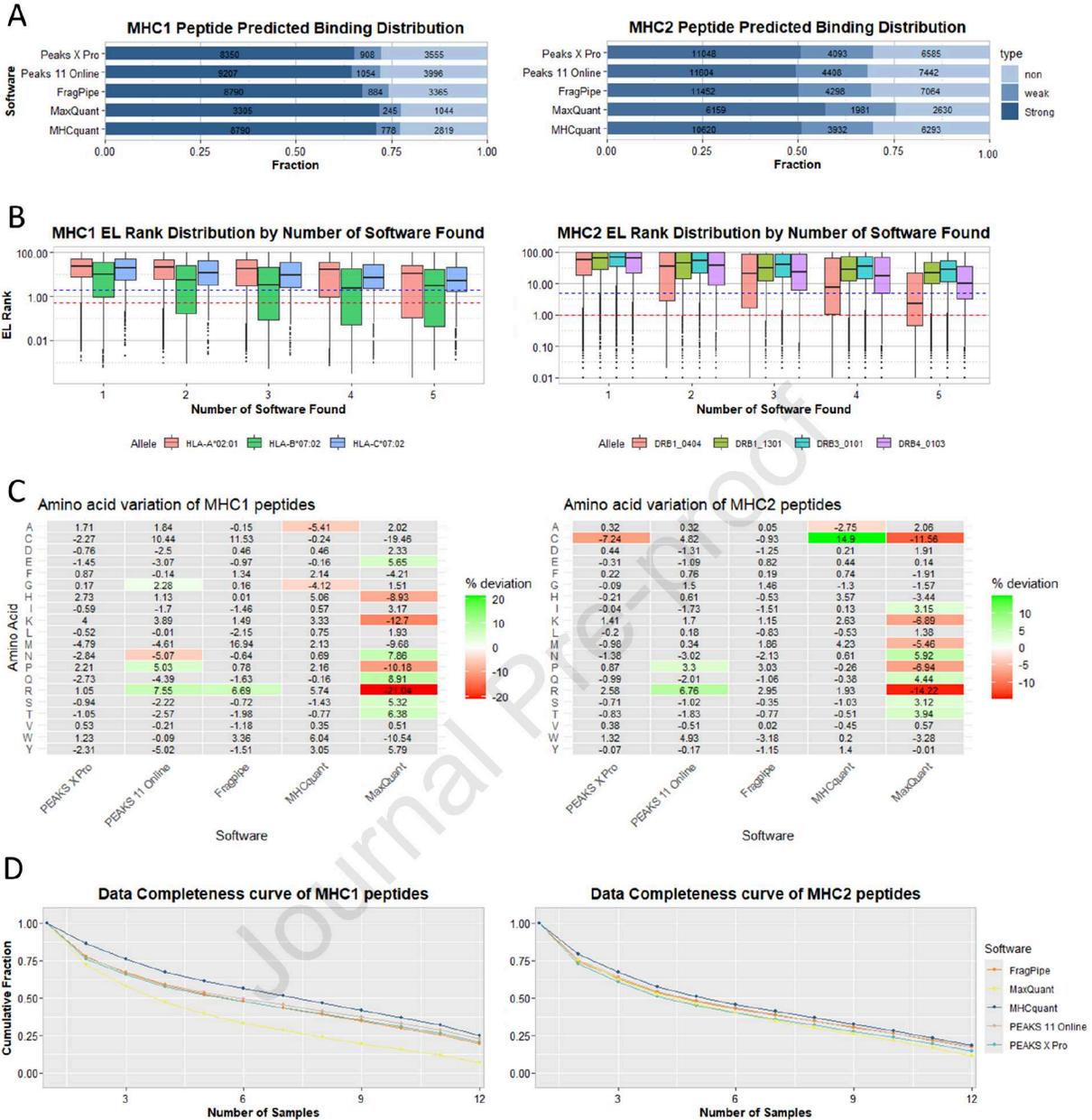
Figure 3. Analysis of identification confidence score between search engine results. A) Pairwise Spearman rank correlation analysis of all software. Bottom left section displays scatterplot of ranks of shared peptides between software (x and y axes correspond to the scores of the software indicated in the top or right header, respectively). Top right section returns the respective spearman rank correlation coefficient. Line plots in the center shows the rank distribution of the data. For FragPipe, hyperscore is used instead of the probability due to the lower precision caused by rounding, which leads to many shared ranks at high probabilities. Since hyperscore is used to calculate probability, they are highly correlated, making hyperscore a suitable alternative. The remaining software uses the standard values for rank calculation (i.e. X.10LgP for PEAKS, PEP for MaxQuant and Score for MHCquant). B) Box and whisker plot showing the relative score rank distribution of peptides found by specific number of software. Lower rank indicates a more confident identification.

Figure 4. Identification performance of MHC1 peptides of different software using decoy and entrapment-based FDR control strategy. Each graph shows the number of target peptide identified at each peptidofom FDR percentage level when using the standard decoy method (red) or the entrapment strategy (blue). Since MHCquant does not report Decoys, the software was run at 1%, 5% and 10% peptidofom FDR to obtain datapoints for comparison and displayed as "X". y-axis limits were set to 75000.

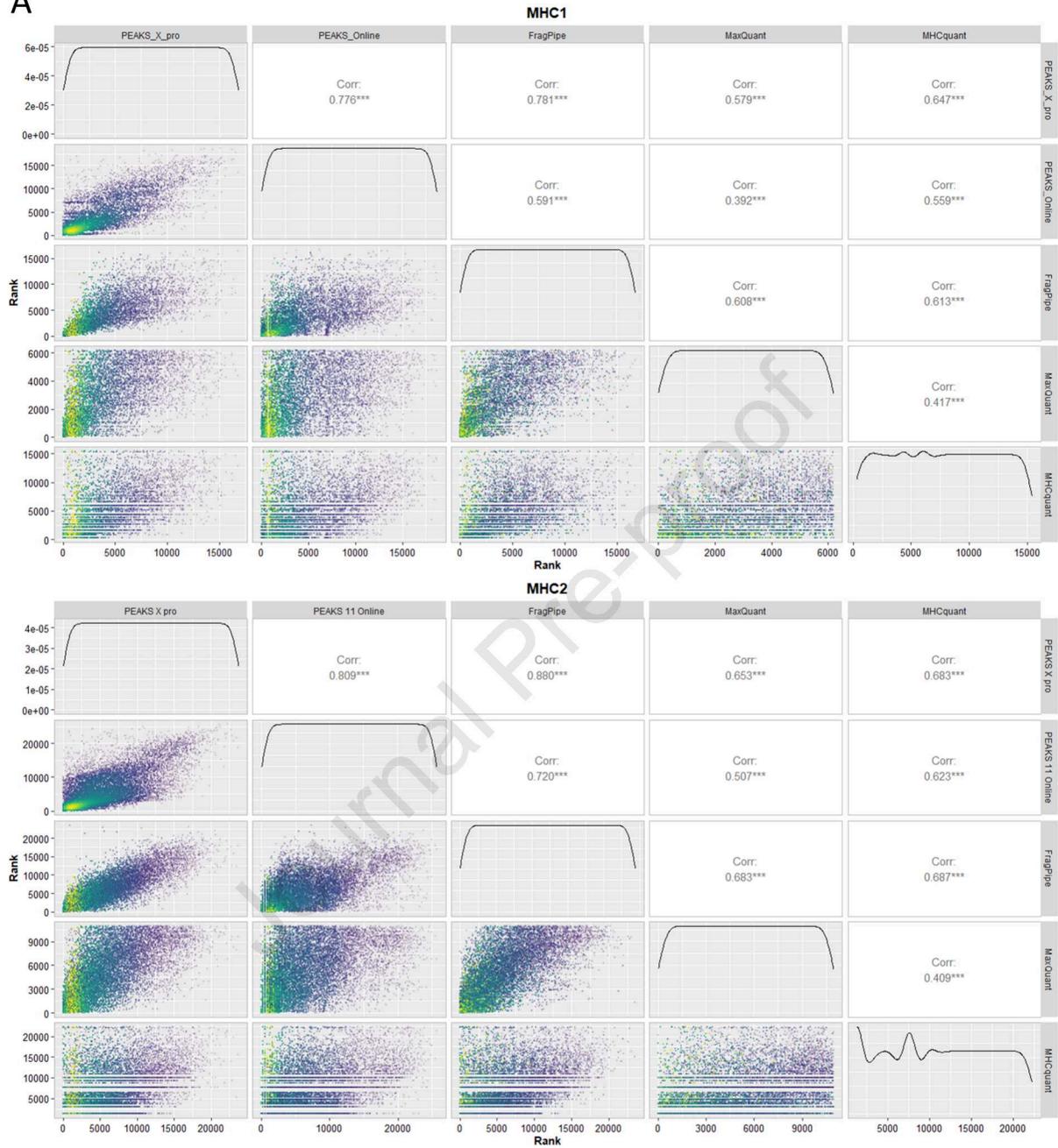
Figure 5. Comparison of the MHC1 identification results from different software given databases of various sizes originating from UniProt reference proteome, RNA-seq transcriptome data and de-novo transcriptome assembly (DNNTA). A) Grouped Stacked bar chart show the total number of identified peptidofom candidates from different software using different databases. Lighter color represents candidates not from UniProt reference database. (For the right plot, the non-UniProt reference fraction has been moved to plot C for clarity). B) Table that shows the complexity of the databases. C) Bar plot shows the number of non-reference identification at 1% FDR. D) Grouped stacked bar plot showing the fraction of non-reference peptide identifications at 1% PSM FDR filtering that are single amino acid variant (SAV) relative to the UniProt reference database and the fraction of these which are I/L variants (substitution of isoleucine with leucine or vice versa.).

Figure 6. Figures showing general properties of the non-reference peptide identifications. A) Upset plot showing the shared identifications. Groups smaller than 10 are omitted. B) Length distribution plot of non-reference peptides. C) Binding prediction of non-reference peptides. D) Charge distribution of non-reference peptides. E) Comparison of the score distribution between the non-reference peptide identification sub-population and the reference peptide identification sub-population. T-test values shown between the compared groups. For Fragpipe, the Hyperscore was used again for better stratification of data. F) Peptide motif of non-reference peptide identifications as reported by MhcVizPipe.

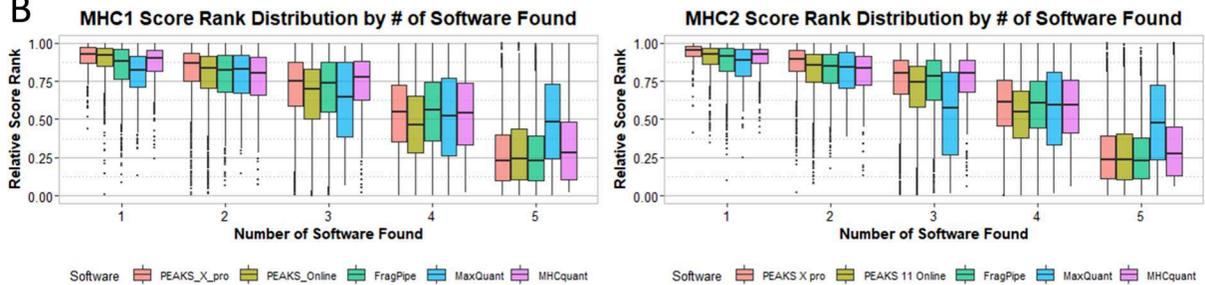


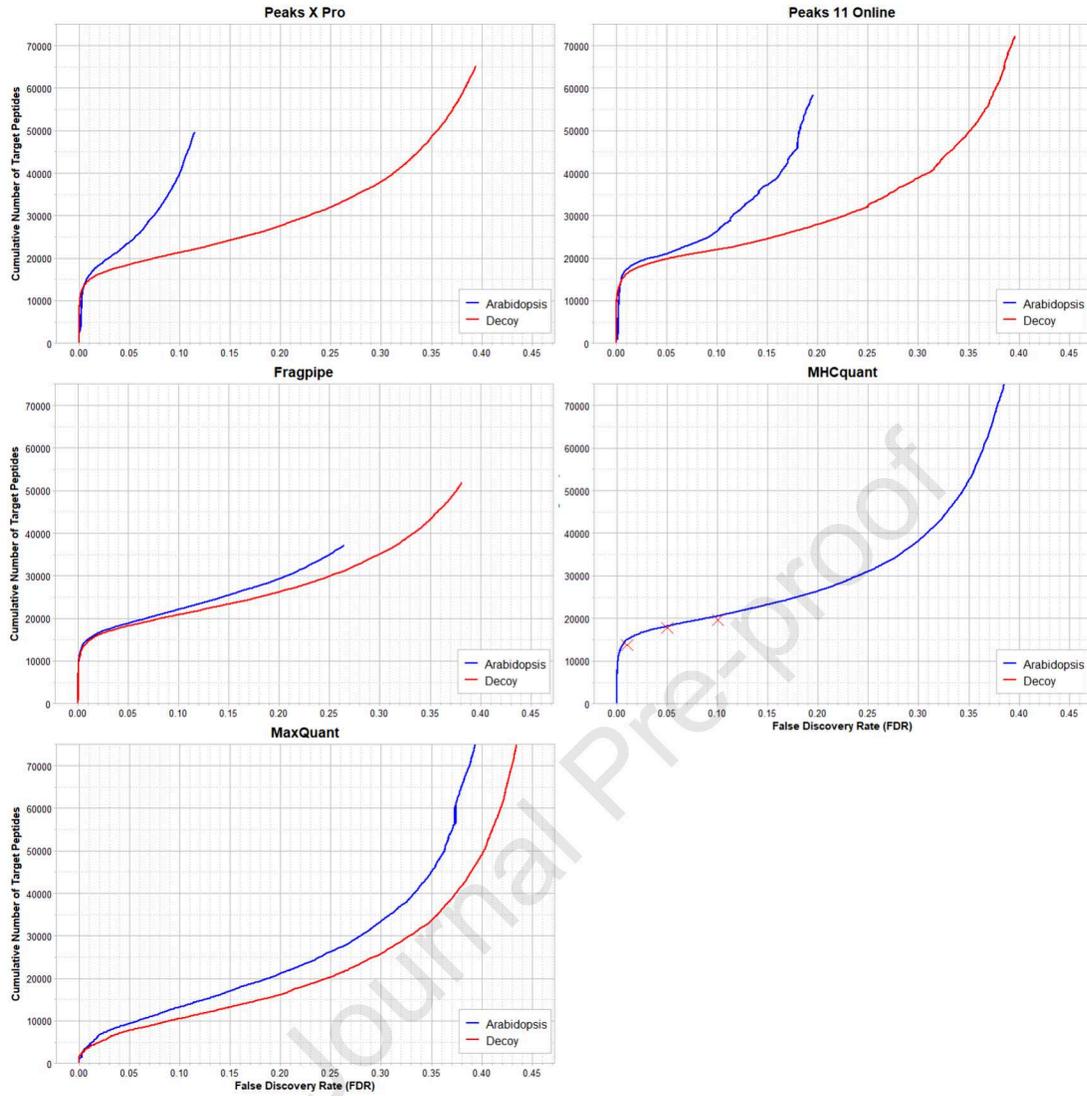


A



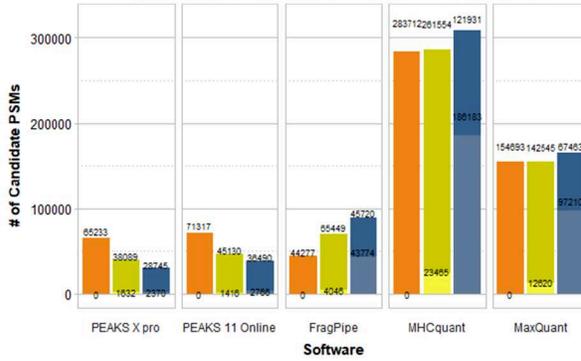
B



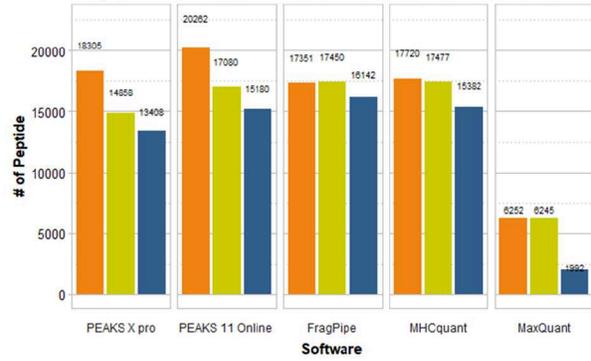


A

Peptidoform Candidate PSMs across Databases



Peptidoform Identifications at 1% FDR across Databases



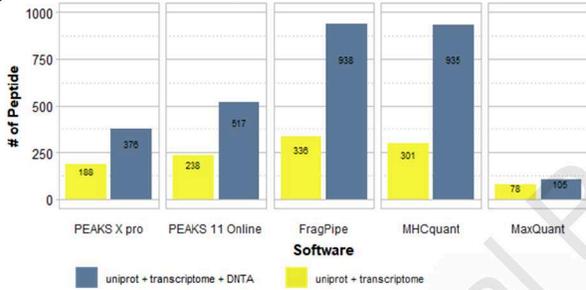
B

Databases	# of unique 9mer (x10 ⁴)
Uniprot	10.34
Uniprot + Transcriptome	11.35
Uniprot + Transcriptome + DNTA	29.59

■ Uniprot ■ Uniprot + transcriptome ■ Uniprot + transcriptome + DNTA
■ reference ■ non-reference

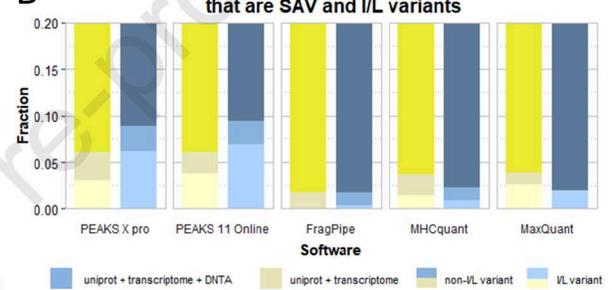
C

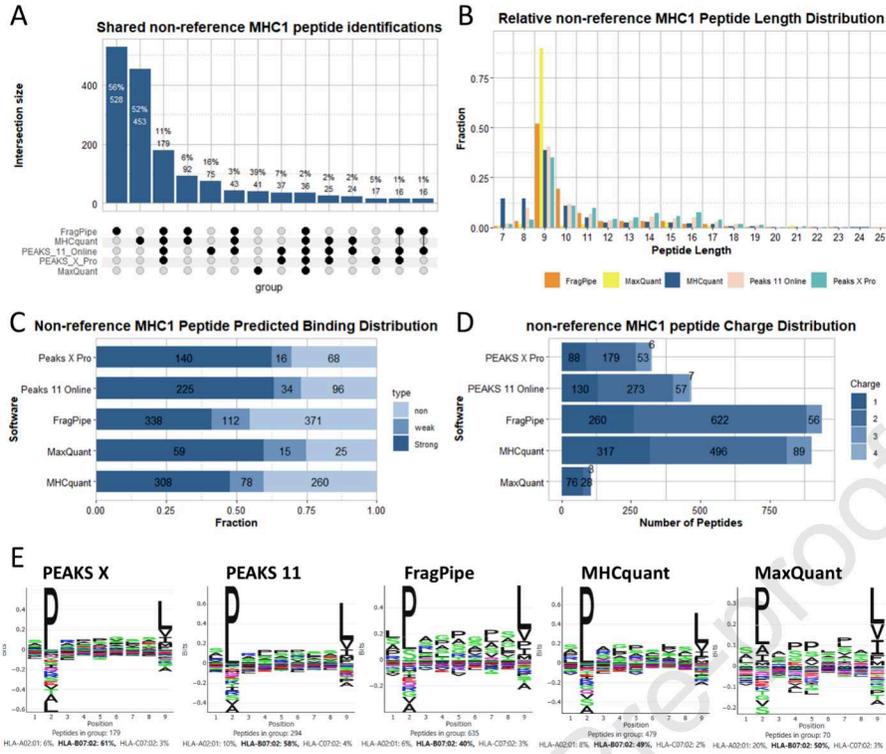
Non-Reference Peptidoform Identifications at 1% FDR



D

Fraction of non-reference identifications that are SAV and I/L variants





Comprehensive benchmarking of database search engines for MHC class 1 & 2 peptide DDA-PASEF data.

Validation of FDR control of search engines through an entrapment-based filtering strategy.

Journal Pre-proof

Large reference database size can have a negative impact on numbers of identification.

PEAKS 11 identified most peptides overall; FragPipe and MHCquant identified most non-reference peptides.

PEAKS, Fragpipe and MHCquant identified similar numbers of non-isobaric single AA variants.

In Brief

Database searching is a key step in mass spectrometry-based proteomics, identifying peptide sequences from raw spectra. Immunopeptidomics poses unique challenges due to the lack of defined proteolytic specificity. We benchmarked leading database search tools: PEAKS (X Pro and 11), FragPipe, MHCquant, and MaxQuant, on DDA-PASEF immunopeptidomics data. PEAKS 11 achieved the highest total identifications, whereas FragPipe and MHCquant detected the most non-reference peptides, highlighting complementary strengths among search engines.

Journal Pre-proof

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The author is an Editorial Board Member/Editor-in-Chief/Associate Editor/Guest Editor for this journal and was not involved in the editorial review or the decision to publish this article.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: