MS²Rescore: Data-driven rescoring dramatically boosts immunopeptide identification rates

Arthur Declercq[1,2], Robbin Bouwmeester[1,2], Sven Degroeve[1,2], Lennart Martens[1,2 §], and Ralf Gabriels[1,2]

[1] VIB-UGent Center for Medical Biotechnology, VIB, Belgium
[2] Department of Biomolecular Medicine, Ghent University, Belgium

[§] To whom correspondence should be addressed
Tel: +32 9 264 93 58
Email: lennart.martens@vib-ugent.be
Address: Technologiepark 75, 9052 Ghent, Belgium

ORCID IDs:

Arthur Declercq: https://orcid.org/0000-0002-9376-1399/
Robbin Bouwmeester: https://orcid.org/0000-0001-6807-7029/
Sven Degroeve: https://orcid.org/0000-0001-8349-3370/
Lennart Martens: https://orcid.org/0000-0003-4277-658X/
Ralf Gabriels: https://orcid.org/0000-0002-1679-1711/

Abbreviations

FDR, false discovery rate

HCD, higher-energy collision-induced dissociation

HLA, human leukocyte antigen

LC-MS, liquid chromatography-mass spectrometry

MS2, tandem mass spectrometry

MHC, major histocompatibility complex

MGF, mascot generic format

USI, universal spectrum identifier

PCC, Pearson correlation coefficient

PSM, peptide-spectrum match

Abstract

Immunopeptidomics aims to identify immunopeptides, which are presented on Major Histocompatibility Complexes (MHC) on every cell and can be used to develop vaccines against pathogens and cancer. However, existing immunopeptidomics data analysis pipelines have some major hurdles to overcome, mostly due to the non-tryptic nature of immunopeptides, which complicates their identification. Previously, the machine and deep learning tools MS²PIP and DeepLC have shown to improve tryptic peptide identifications by accurately predicting tandem mass spectrometry (MS2) peak intensities and retention times, respectively, and by using these predictions to rescore peptide-spectrum matches (PSMs) with post-processing tool Percolator. However, MS²PIP was still tailored towards tryptic peptides and fragmentation patterns of immunopeptides are drastically different. To enable MS²PIP-based rescoring of immunopeptide PSMs, we have retrained MS²PIP to include non-tryptic peptides. These newly trained MS²PIP models greatly improve the predictions for immunopeptides and, surprisingly, also for tryptic peptides. Next, the new MS²PIP models, DeepLC, and Percolator were integrated into one software package, called MS²Rescore. Using MS²Rescore, 46% more spectra and 36% more unique peptides were identified at 1% false discovery rate (FDR), with even more extreme differences at 0.1% FDR, in comparison with standard Percolator rescoring. Due to the innovative extraction of MS²PIP-, DeepLC and search engine-based features, MS²Rescore even outperforms current state-of-the-art immunopeptide rescoring efforts. Thus, the integration of the new immunopeptide MS²PIP models, DeepLC, and Percolator into MS²Rescore shows great promise to substantially improve the identification of novel epitopes from immunopeptidomics workflows.

Introduction

The immune system is a complex, yet remarkable system that protects us from both invaders from outside the body, i.e. pathogens, as well as from inside the body, i.e. malignancies (1). Increased understanding of the immune system allowed for great medical achievements such as vaccination, which is currently available for over 29 diseases, enabled the eradication of smallpox, and prevents over 3 million deaths each year (2). However, many diseases such as *Mycobacterium tuberculosis* or malignancies lack effective vaccines due to improper T-cell activation. A key issue in developing effective vaccines for these diseases is the lack of accurately identified major histocompatibility complex (MHC)-presented epitopes or immunopeptides. These epitopes are presented on the cell surface and enable T-cells to discern healthy cells from infected or malignant cells. While much effort has recently been invested in accurate prediction of these epitopes *in silico* (3), these are mostly limited to viruses as these contain fewer potential protein antigens (4). Moreover, these tools are not yet sufficiently precise to confidently identify epitopes (5, 6). Therefore, experimental immunopeptidomics workflows, such as epitope detection through liquid chromatography-mass spectrometry (LC-MS), are still the best way to accurately identify these immunopeptides (7).

While immunopeptidomics workflows have been readily developed and applied (8), acquisition of immunopeptides through LC-MS suffers some major problems. First, the acquisition of immunopeptide spectra is hampered due to the low abundance of immunopeptides and even more so of neo-epitopes. Infrequently occurring epitopes are still very challenging to identify through LC-MS, despite enrichment efforts and sample preprocessing (9). Second, in contrast to standard proteomics experiments where proteins are usually digested with trypsin before LC-MS, immunopeptides are acquired through immune purification with antibodies followed by acidic elution resulting in mostly non-tryptic peptides. The non-tryptic nature of immunopeptides results in one less positive charge due to the missing arginine or lysine at the end of the peptide, causing many immunopeptides to be singly charged during MS acquisition. These singly charged peptides are much harder to analyze because, after fragmentation of the peptide, the charge resides on one of the fragments, leaving the other uncharged and thus lost (10). Moreover, most contaminants are singly charged as well, making identifications of immunopeptides much harder (11). The non-tryptic nature of immunopeptides hampers not only the acquisition but also the identification of immunopeptide spectra. To match each acquired spectrum with the peptide from which the spectrum originated, proteomics database search engines such as SEQUEST (12), X!Tandem (13), Andromeda (14), or PEAKS (15) generate *in silico* spectra for all potentially matching peptides. The complete list of peptides that could be in the sample is called the search space. It is important to note that spectra from peptides that are not included in this search space cannot be identified,

95 even though they were acquired. In standard shotgun proteomics experiments, *in silico* tryptic
96 digestion of relevant proteins yields a broad yet representable search space. In
97 immunopeptidomics, however, the search space tends to be two orders of magnitude larger due
98 to (i) seemingly random splicing from the protein of origin, (ii) the variable length of MHC class I
99 bound peptides, 8-11 amino acids, and MHC-II peptides, 6-24 amino acids (16) and (iii) the potential
100 occurrence of conformational cis- and trans-spliced immunopeptides, which are non-linear
101 peptides that originate from the same or different proteins, respectively (17). This search space
102 expansion results in considerably more ambiguity between candidate peptide-spectrum matches
103 (PSMs) (18), lower PSM scores, drastically elevated false discovery rate (FDR) score thresholds, and
104 ultimately in fewer identified immunopeptides (19). Furthermore, because tryptic peptides have
105 been the longtime standard in proteomics, search engines as well as bioinformatics tools that aid
106 in identifying LC-MS spectra, are tailored towards tryptic peptides, making them less accurate or
107 not applicable at all for immunopeptidomics.

108 The high need for neo- and xeno-epitope discovery led to the development of many bioinformatics
109 tools to improve or validate identifications in immunopeptidomics. On the one hand, motif
110 deconvolution tools have been developed that leverage binding motifs of immunopeptides to
111 validate immunopeptide identifications. On the other hand, full pipelines have been developed to
112 improve immunopeptide identification. For example, MHCquant (20), which is a recent
113 computational workflow designed specifically for neo-epitope identification, and PEAKS DB (15).
114 Even though PEAKS DB is not specifically designed for immunopeptides, it is highly interesting due
115 to its *de novo* assisted database searches, which tend to work well for large search spaces. Even
116 though these tools can help with immunopeptide identification, they do not use all available
117 information, such as retention time and fragment ion intensity patterns. Previously, it has been
118 proven that integrating retention time predictions in standard proteomics workflows can improve
119 identification rates (21). Similarly, adding peak intensity predictions to post-processing tools such
120 as Percolator can also improve identification rates drastically (22), which has already been proven
121 to work for immunopeptides as well by efforts such as Prosit (23, 24). Similarly, tools such as
122 DeepLC (25) and MS²PIP (26–28) can provide accurate retention time predictions and peak intensity
123 predictions, respectively, to aid in post-processing. Indeed, when combined with Percolator,
124 identification rates at a fixed FDR have been proven to substantially increase (22). However,
125 currently, DeepLC and MS²PIP are solely trained on tryptic peptides. This absence of lysine and
126 arginine at the C-terminus is less of a problem for DeepLC as the effect on retention time is small
127 and is accounted for through feature encoding (29). However, this is not the case for MS²PIP, as
128 alterations in peptide composition, as well as fragmentation patterns and labeling methods heavily

129 alter peak intensity patterns (26). Therefore, we here present greatly improved MS²PIP models that
130 include immunopeptides, and non-tryptic peptides in general. Moreover, we have integrated
131 MS²PIP and DeepLC with Percolator into the free and open source MS²Rescore software tool, which
132 enables improved rescoring of peptide identifications from various proteomics search engines.
133 Altogether, we show that well-adapted fragmentation spectrum and retention time predictions
134 integrated into MS²Rescore drastically increase immunopeptide identification rates and
135 outperform existing post-processing methods.

Experimental Procedures

To train and test new MS²PIP models, five publicly available immunopeptide data sets and one publicly available chymotrypsin-digestion data set were downloaded from PRIDE Archive (30, 31). Similarly, for evaluation on representative unseen data, four distinct data sets were downloaded: (i) a data set containing HLA-I immunopeptides, (ii) a data set containing HLA-II immunopeptides, (iii) the data set with tryptic peptides that was previously used to evaluate the existing MS²PIP HCD models, and (iv) a data set containing chymotrypsin-digested peptide data. The corresponding ProteomeXchange project identifiers for each data set are listed in Supplementary Table S1.

As tandem mass spectrometry (MS2) fragmentation patterns are highly dependent on the instrument, instrument settings, fragmentation method, and any applied labeling methods (26), all MS²PIP train, test, and evaluation data must originate from experiments with the same experimental parameters. Unlabeled higher-energy collision-induced dissociation (HCD) data from Orbitrap instruments was used, as this makes the newly trained models widely applicable and plenty of training data is readily available on public repositories. For each PRIDE Archive project, the raw mass spectrometry files were converted to Mascot Generic Format (MGF) files using ThermoRawFileParser (v1.3.4) (32). The corresponding identification files were converted to MS²PIP input files using custom Python scripts and were further filtered to retain unique combinations of peptide sequence, modifications, and precursor charge at 1% FDR. Next, all spectra were combined into one MGF file. The universal spectrum identifier (USI) (33) was used as unique identifier for each PSM to ensure reproducibility and a one-on-one mapping between peptide identifications and spectra. Data from each PRIDE Archive project was either used as train/test data or as evaluation data, except for the chymotrypsin data, where the same project was used to provide both training/testing data (70%) and evaluation data (30%). This split was made after selecting unique peptide-modification-charge combinations, to ensure no overlap in samples between both splits.

Similarly to the 2019 MS²PIP models, new models were trained with the XGBoost machine learning algorithm (34). The Hyperopt (35) package (v0.2.5) was used in combination with four-fold cross-validation for hyperparameter optimization, allowing 400 boosting rounds and early stopping fixed at 10 rounds. Hyperparameters were optimized for each training data set separately, as well as for b- and y-ion models. All selected hyperparameters are shown in Supplementary Table S2. To evaluate each model, the Pearson correlation coefficient (PCC) was calculated between observed and predicted b- and y-ion peaks for each spectrum. The model performances were further analyzed by peptide length and precursor charge. Ultimately, three models were trained for each ion type: the *immunopeptide* model solely trained on immunopeptides, the *immunopeptide-*

169 *chymotrypsin* model trained on immunopeptides supplemented with chymotrypsin-digested
170 peptides, and the *non-tryptic immunopeptide model* solely trained on non-tryptic
171 immunopeptides. Ultimately, two models were integrated into MS²PIP: (i) the *immunopeptide*
172 model and (ii) the *immuno-chymotrypsin* model. The former can be used for immunopeptide peak
173 intensity predictions and the latter for tryptic and more general non-tryptic peptide predictions.
174 For further analysis into rescoring immunopeptide PSMs, only the *immunopeptide* model was
175 used.

176 To validate the capacity of the new MS²PIP models to improve immunopeptide identification rates,
177 the new models were implemented with DeepLC (v0.1.36) and Percolator (v3.5) into MS²Rescore.
178 MS²Rescore calculates various meaningful features based on (i) the search engine output, (ii) the
179 DeepLC-predicted and the observed retention times, and (iii) the MS²PIP-predicted and the
180 observed MS2 peak intensities. These features are then passed to Percolator for PSM rescoring.
181 Search engine features were selected based on the previous publication by Granholm et al. (36)
182 and replicated for use with MaxQuant search results (37). MS²PIP features were used as first
183 described by Silva et al (22). All features generated by MS²Rescore are listed in Supplementary
184 Table S3. MS²Rescore was validated on a large-scale immunopeptide data set (38) which was also
185 used to validate the recently published Prosit rescoring effort for immunopeptides (PXD021398)
186 (24). This allows both an evaluation of the improved identification rates due to the new MS²PIP
187 models, and a straight-forward comparison with Prosit rescoring. First, the msms.txt identification
188 files for the projects' two MaxQuant searches (alkylated and non-alkylated samples), the Prosit-
189 rescored Percolator output files, and the raw mass spectrometry files were downloaded from
190 PRIDE Archive. The mass spectrometry files were then further processed with
191 ThermoRawFileParser (v1.3.4) (32) and the PSMs for each of the two MaxQuant searches were
192 rescored separately. Two rescoring methods were evaluated: (i) Using only search engine features,
193 replicating a normal Percolator run, and (ii) using the full MS²Rescore feature set, including search
194 engine-, MS²PIP-, and DeepLC-features. Additionally, these rescoring methods were compared with
195 the original MaxQuant results and the downloaded Prosit-rescoring results.

196 Finally, each method was evaluated at varying FDR thresholds in terms of identification rate and
197 number of unique identified peptides. The contribution of the different feature sets in MS²Rescore
198 was visualized using Percolator's model weights, and the distributions of retention time difference
199 and MS²PIP prediction correlations were compared between decoy PSMs, accepted target PSMs,
200 and rejected target PSMs. All plots were generated in Jupyter notebooks (v6.4.0) using Python
201 (v3.8.3) with the Matplotlib (v3.4.2) (39), Seaborn (v0.11.0) (40) , and spectrum_utils (v0.3.5) (41)
202 libraries.

Results

In order to improve the identification rate of immunopeptides by leveraging peak intensity predictions, new models for MS²PIP were trained for immunopeptides. Despite using different training set compositions, all newly trained models drastically improve predictions for both HLA-I and HLA-II data in comparison with the tryptic 2019 HCD model (Figure 1A). Surprisingly, even for standard tryptic shotgun proteomics data the predictions from the new models are slightly better, largely due to the portion of tryptic peptides within the immunopeptide training data. Indeed, when these peptides are left out of the training data, accuracy drops in comparison with the 2019 HCD model. While both immunopeptide models are well suited to predict peak intensities for tryptic- and immunopeptides, the performance on chymotrypsin-digested peptides is not as high. Only when chymotrypsin-digested peptides are included in the training data, performance on chymotrypsin data increases. This difference in performance between models trained with or without chymotrypsin-digested peptides, is mainly seen for longer peptides and peptides with higher precursor charge states (Supplementary Figure S1). As models that are solely trained on immunopeptides have not seen longer peptides with higher charge-states, these models can not accurately model how these will behave in the mass spectrometer. Thus, even though immune- and chymotrypsin-digested peptides are both considered non-tryptic, they are still very different. Overall, immunopeptide peak intensity predictions are drastically improved by all the newly trained models, with the immunopeptide model showing the highest accuracy (median PCC of 0.96). The exact median PCCs are listed in Supplementary Table S4. Examples of a prediction with median PCC values with the immunopeptide model and the corresponding, less accurate, 2019 model prediction are shown in Figure 1B and Figure 1C.

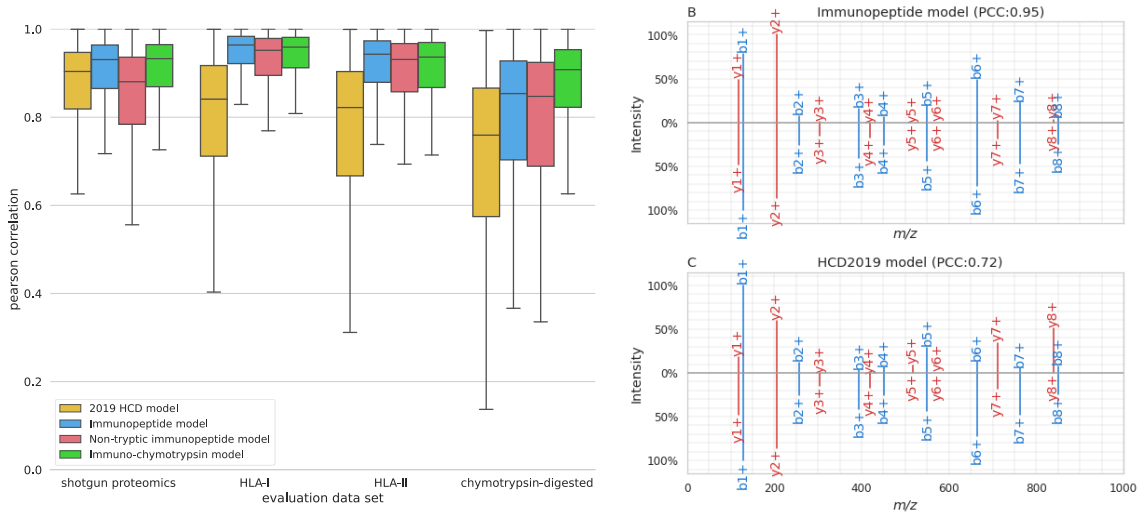


Figure 1: (A) Boxplots comparing Pearson correlation distributions between predicted and observed peak intensity values on spectrum level for the 2019 HCD model, immunopeptide model, non-tryptic immunopeptide model, and immuno-chymotrypsin model. Performance is evaluated on general tryptic proteomics data, HLA-I immunopeptide data , HLA-II immunopeptide data , and chymotrypsin data. (B) Predictions for immunopeptide "KQHGVNVSV" by new immunopeptide MS²PIP model (top) compared to observed MS2 spectrum (mzspec:PXD005231:20160513_TIL1_R1:scan:10909) (bottom). This peptide was selected for visualization as its PCC lies close to the median for the immunopeptide model on all HLA-I evaluation data. (C) Predictions by the 2019 HCD model for the same immunopeptide as visualized in B.

Ultimately, the goal of these newly trained models is to improve immunopeptide identification rates by providing more accurate peak intensity predictions for immunopeptides. Therefore, (i) identification results without rescoring, (ii) rescoring with solely search engine features (replicating a normal Percolator run), (iii) rescoring with MS²Rescore (including DeepLC and the new MS²PIP models), and (iv) rescoring with the recently published Prosit models were compared in terms of the total amount of identifications as well as the number of unique identifications based on sequence. Overall, rescoring with both MS²Rescore and Prosit substantially improved the spectrum identification rate in comparison with rescoring with search engine features alone or not rescoring, and this at both 1% and 0.1% FDR. Indeed, MS²Rescore achieves an identification rate of 11,1%, out of 18 million spectra, which is 46% more compared to 7.6% for traditional rescoring, and only 1.9% for the MaxQuant search results, all at 1% FDR (Figure 2A-B). Thus, providing peak intensities and retention time predictions to Percolator substantially increases the number of identified immunopeptides. This is clearly illustrated by analyzing the Percolator weights for each separate feature as well as the combined absolute weights for search engine features, MS²PIP features and DeepLC features (Supplementary Figure S2). Similarly, the number of unique identified immunopeptides increases by 36% when adding MS²PIP and DeepLC Features for the 1% FDR, and even more so for 0.1% FDR where the number of unique identified peptides reaches nearly 300% of the number of traditional Percolator identification results (Figure 2C-D, Supplementary FigureS3A-B).

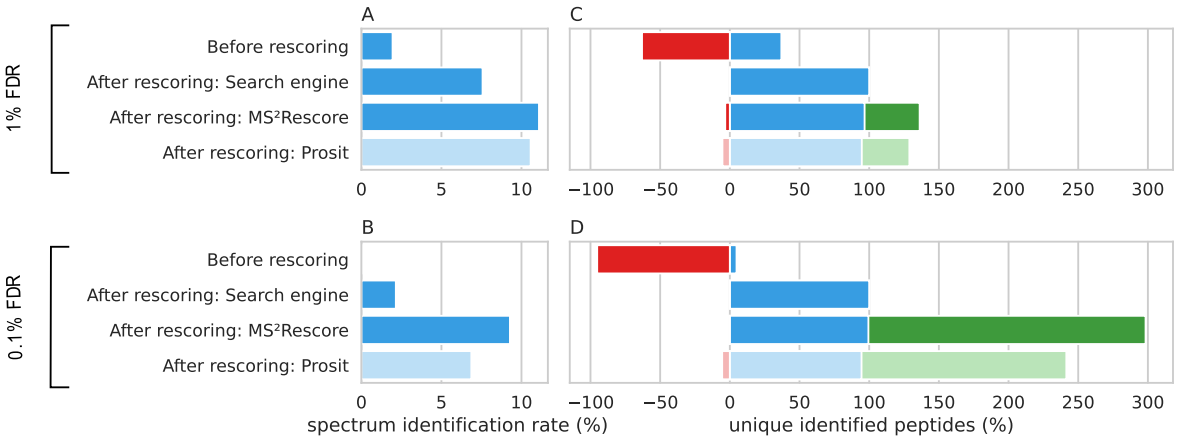


Figure 2: (A) Bar chart showing the spectrum identification rate at 1% FDR out of 18.375.659 spectra. (B) Bar chart showing the spectrum identification rate at 0.1% FDR out of 18.375.659 spectra. (C) Bar chart with the shared (blue), gained (green) and lost (red) number of unique (by sequence) identified immunopeptides in relation to rescoring with only search engine features for the 1% FDR threshold. (D) Bar chart with the shared (blue), gained (green) and lost (red) number of uniquely (by sequence) identified immunopeptides in relation to rescoring with only search engine features for the 0.1% FDR threshold.

The power of providing these predictions to Percolator is further illustrated when visualizing the distributions for decoy PSMs, rejected target PSMs, and accepted target PSMs. Indeed, the distributions for decoy and rejected target PSMs are highly similar for both the retention time error as well as the PCC, while the accepted target PSMs accumulate around low retention time errors and high PCCs (Figure 3A-B). The accepted target PSMs are clearly separable from the decoy and rejected target PSMs using only the PCC and retention time error distributions (Figure 3C-D). Furthermore, while both metrics correlate with the search engine score, a large amount of decoy and rejected target PSMs can only be separated from the target PSMs by also including PCC or retention time error information (Figure 3E-F). This clearly illustrates how Percolator achieves its much improved separation between good and bad target PSMs when provided with peak intensity and retention time prediction features. Furthermore, PSMs that previously would have been incorrectly accepted below a 1% FDR because of a high search engine score alone, are now rejected due to a low PCC, a high retention time error, or both. This most likely accounts for the small percentage of identified peptides that are lost after rescoring.

The integration of MS²PIP-, DeepLC-, and search engine-based features in MS²Rescore has proven to substantially increase the identification rate of immunopeptides, and furthermore, performs better than the recently published Prosit rescoring method (24). In comparison with Prosit, MS²Rescore gains 5% and 35% more identifications at 1% FDR and 0.1% FDR, respectively. This trend continues for the number of unique identified peptides with a respective increase of 8% and 57% (Figure 2). MS²Rescore thus substantially increases the identification rate, especially for more

stringent FDR thresholds. Peptides for which Prosit cannot predict MS2 peak intensities, i.e. unmodified Cys, cysteinylated Cys, and acetylated peptide N-termini, as well as other PSMs, were left out of the Prosit rescoring output. That is why MS²Rescore ultimately ends up with more identifications in the unfiltered dataset, i.e. 100% FDR (Supplementary Figure S4A-B), especially for the noIAA sample. For a thorough comparison, these were left out of the MS²Rescore output as well. However, this seemed to have a rather negligible impact on the number of identifications at 1% and 0.1% FDR (Supplementary Figure S4C-D), and thus the difference in identification rates cannot be attributed to these peptides. Altogether, the improved immunopeptide peak intensity predictions from the new immunopeptide MS²PIP model, combined with retention time predictions of DeepLC drastically improves the identification rates of immunopeptides and highly increases the confidence in the accepted PSMs.

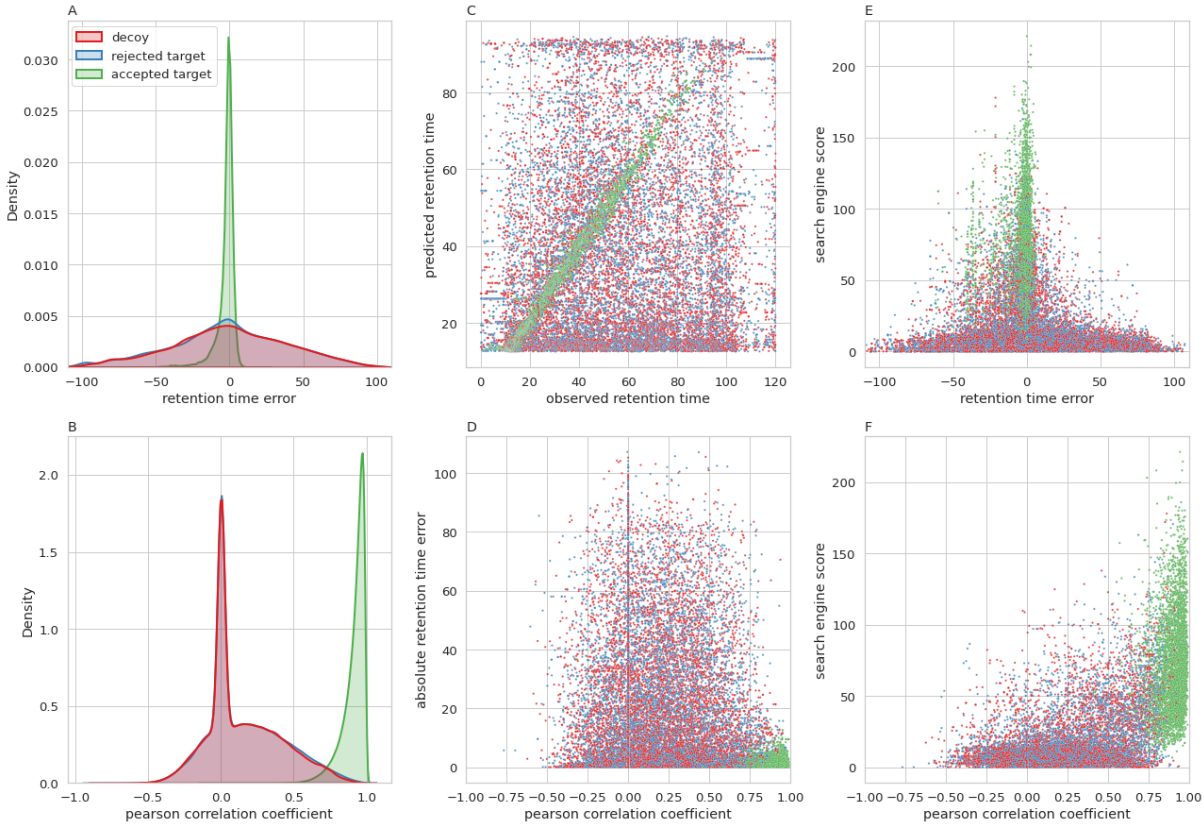


Figure 3 (A) Density plot showing the distribution of the smallest retention time error between observed and predicted retention time for each unique combination of sequence and modification for decoys (red) rejected targets ,q value > 0.01(blue) and accepted targets, q value < 0.01 (green).(B) Density plot showing the Pearson correlation between observed and predicted peak intensities (note the rejected targets distribution coincides with the decoy distribution). Scatterplot showing the relation (C) between observed and predicted retention time, (D) between Pearson correlation and retention time error, (E) between retention time error and search engine score, and (F) between Pearson correlation and search engine score for decoys (red), rejected targets (blue), and accepted targets (green) for all PSMs from "GN20170531_SK_HLA_C0102_R1_01". Note that all retention time errors are the smallest retention time error for each precursor (see rt_diff_best table S3), all Pearson correlation coefficients are calculated for each full log2 transformed spectrum, and all zero values are caused by the fact that either the observed or predicted intensities for a given ion type are all zero.

Discussion

By training new peak intensity prediction models, we were able to greatly enhance immunopeptide identification rate through PSM rescoring. While all newly trained models greatly enhance peak intensity predictions for immunopeptides, the model trained solely on immunopeptides performed best. Even though the immuno-chymotrypsin model contained the same immunopeptide train set, the addition of the chymotrypsin-digested peptides did lower the performance slightly. Similarly, not including chymotrypsin-digested peptides in the training data resulted in lower accuracies for the chymotrypsin-digested peptides. Indeed, immunopeptides are generally much smaller and consequently carry a lower charge state. Therefore, immunopeptide trained models are not able to predict the behavior of longer and higher charged peptides in the mass spectrometer. While both immuno- and chymotrypsin-digested peptides are considered non-tryptic, their properties can be very different, leading to reduced accuracy of peak intensity prediction models when applied on a different type of non-tryptic peptides. Therefore, a general non-tryptic peak intensity prediction model seems not ideal, as it could lower performance for subsets like immunopeptides. Surprisingly, while immuno- and chymotrypsin-digested peptides are antagonistic, immuno- and tryptic peptides seem synergistic in terms of training data. This comes as no surprise as almost 50% of the immunopeptide training data consists of tryptic peptides. Immunopeptides are thus not necessarily non-tryptic. However, the actual occurrence of tryptic peptides in immunopeptidomics samples is most likely much lower. This unrepresentatively sized tryptic portion most likely originated from the tryptic bias in current immunopeptidomics workflows. Indeed, in previous studies, tryptic MHC peptide coverage could rise to 70% (42). By training new, non-tryptic models of MS²PIP, we take a first step in decreasing this tryptic bias, to ultimately be able to analyze an unbiased immunopeptide landscape.

Moreover, by integrating the new immunopeptide model with retention time predictions and search engine features into MS²Rescore, we greatly enhanced the ability of Percolator to rescore immunopeptide PSMs, resulting in a much improved immunopeptide identification workflow. Furthermore, rescoring drastically increases the number of unique identified peptides, which is of crucial importance for the discovery of potential neo-epitopes for cancer vaccination or xeno-epitopes for anti-bacterial, and to a lesser extent, anti-viral vaccines. Moreover, while previously almost no identifications were found at a more confident 0.1% FDR threshold, MS²Rescore allows a lowering of the FDR threshold to 0.1%, while retaining 83% of the peptides identified at 1% FDR. This illustrates the large increase in confidence of the identified PSMs MS²Rescore allows for.

Additionally, MS²Rescore with DeepLC and the new immunopeptide MS²PIP model shows an improved identification rate over the recently published Prosit effort, especially for lower FDR thresholds. As Prosit has shown to provide more accurate predictions compared to previous MS²PIP models, it is unlikely that MS²Rescore's higher performance can be attributed to superior peak intensity predictions. Most likely the difference in rescoring performance is the result of the generation of more relevant MS²PIP-, DeepLC and search engine-derived features.

MS²Rescore is freely available under the permissive Apache 2.0 open-source license on GitHub (https://github.com/compomics/ms2rescore) and can easily be installed and used locally through the PyPI Python package. Various identification files from different search engines are accepted, and both MS²PIP and DeepLC can handle a variety of modifications, eliminating the need to filter identification files before rescoring. Altogether, these new models show great promise to greatly extend the immunopeptide landscape in existing and future immunopeptidomics experiments.

Acknowledgments

R.B. acknowledges funding from the Vlaams Agentschap Innoveren en Ondernemen under project number HBC.2020.2205.; R.G. received funding from the Research Foundation Flanders (FWO) [1S50918N]. S.D. and L.M. acknowledge funding from the European Union's Horizon 2020 Programme (H2020-INFRAIA-2018-1) [823839]; L.M. acknowledges funding from the Research Foundation Flanders (FWO) [G028821N] and from Ghent University Concerted Research Action [BOF21/GOA/033]. A.D. received funding from the Research Foundation Flanders (FWO) [1SE3722]

Data availability

MS²Rescore is available at https://github.com/compomics/ms2rescore. All additional code used in this work is available at compomics/ms2rescore-immunopeptidomics-manuscript (github.com). The data used for training and evaluation of the newly trained models, the models themselves as well as the MS²Rescore output is available on Zenodo at https://doi.org/10.5281/zenodo.5615050.

References

1. Sattler, S. (2017) in *Advances in Experimental Medicine and Biology* (Springer New York LLC), pp 3–14.

2. Centers For Disease Control And Prevention (2021) Recommended Vaccines By Disease.

3. Raoufi, E., Hemmati, M., Eftekhari, S., Khaksaran, K., Mahmodi, Z., Farajollahi, M. M., and Mohsenzadegan, M. (2020) Epitope Prediction by Novel Immunoinformatics Approach: A State-of-the-art Review. *Int. J. Pept. Res. Ther.* 26, 1155–1163

4. Mayer, R. L., and Impens, F. (2021) Immunopeptidomics for next-generation bacterial vaccine development. *Trends Microbiol.*,

5. Larsen, M. V, Lundegaard, C., Lamberth, K., Buus, S., Lund, O., and Nielsen, M. (2007) Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinforma. 2007 81* 8, 1–12

6. Zhang, H., Lundegaard, C., and Nielsen, M. (2009) Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods. *Bioinforma. Orig. Pap.* 25, 83–89

7. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J., and Mann, M. (2015) Mass spectrometry of human leukocyte antigen class i peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics* 14, 658–673

8. Solleder, M., Guillaume, P., Racle, J., Michaux, J., Pak, H. S., Müller, M., Coukos, G., Bassani-Sternberg, M., and Gfeller, D. (2020) Mass spectrometry based immunopeptidomics leads to robust predictions of phosphorylated HLA class I ligands. *Mol. Cell. Proteomics* 19, 390–404

9. Faridi, P., Purcell, A. W., and Croft, N. P. (2018) In Immunopeptidomics We Need a Sniper Instead of a Shotgun. *Proteomics* 18,

10. Pfammatter, S., Bonneil, E., Lanoix, J., Vincent, K., Hardy, M.-P. P., Courcelles, M., Perreault, C., and Thibault, P. (2020) Extending the Comprehensiveness of Immunopeptidome Analyses Using Isobaric Peptide Labeling. *Anal. Chem.* 92, 9194–9204

11. Purcell, A. W., Ramarathinam, S. H., and Ternette, N. (2019) Mass spectrometry–based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc.* 14, 1687–1707

12. Eng, J. K., McCormack, A. L., and Yates, J. R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5, 976–989

13. Craig, R., and Beavis, R. C. (2004) TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* 20, 1466–1467

14. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R. A., Olsen, J. V., and Mann, M. (2011) Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805

15. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G. A., and Ma, B. (2012) PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* 11,

16. Jiang, J., Natarajan, K., and Margulies, D. H. (2019) in *Advances in Experimental Medicine and Biology* (Springer New York LLC), pp 21–62.

17. Faridi, P., Li, C., Ramarathinam, S. H., Vivian, J. P., Illing, P. T., Mifsud, N. A., Ayala, R., Song, J., Gearing, L. J., Hertzog, P. J., Ternette, N., Rossjohn, J., Croft, N. P., and Purcell, A. W. (2018) A subset of HLA-I peptides are not genomically templated: Evidence for cis- and trans-spliced peptide ligands. *Sci. Immunol.* 3,

18. N, C., S, D., K, H., and L, M. (2011) Analysis of the resolution limitations of peptide identification algorithms. *J. Proteome Res.* 10, 5555–5561

19. Verheggen, K., Ræder, H., Berven, F. S., Martens, L., Barsnes, H., Vaudel, M., K, V., H, R., FS, B., L, M., H, B., and M, V. (2020) *Anatomy and evolution of database search engines—a central component of mass spectrometry based proteomic workflows* (John Wiley & Sons, Ltd)

20. Bichmann, L., Nelde, A., Ghosh, M., Heumos, L., Mohr, C., Peltzer, A., Kuchenbecker, L., Sachsenberg, T., Walz, J. S., Stevanović, S., Rammensee, H. G., and Kohlbacher, O. (2019) MHCquant: Automated and Reproducible Data Analysis for Immunopeptidomics. *J. Proteome Res.* 18, 3876–3884

411 21. Dorfer, V., Maltsev, S., Winkler, S., and Mechtler, K. (2018) CharmeRT: Boosting Peptide Identifications
412 by Chimeric Spectra Identification and Retention Time Prediction. *J. Proteome Res.* 17, 2581–2589

413 22. C Silva, A. S., Bouwmeester, R., Martens, L., Degroeve, S., C. Silva, A. S., Bouwmeester, R., Martens, L.,
414 Degroeve, S., Silva, A. S. C., Bouwmeester, R., Martens, L., Degroeve, S., and Wren, J. (2019) Accurate
415 peptide fragmentation predictions allow data driven approaches to replace and improve upon
416 proteomics search engine scoring functions. *Bioinformatics* 35, 1401–1403

417 23. Li, K., Jain, A., Malovannaya, A., Wen, B., and Zhang, B. (2020) DeepRescore: Leveraging Deep Learning
418 to Improve Peptide Identification in Immunopeptidomics. *Proteomics* 20,

419 24. Wilhelm, M., Zolg, D. P., Graber, M., Gessulat, S., Schmidt, T., Schnatbaum, K., Schwencke-Westphal, C.,
420 Seifert, P., Krätzig, N. de A., Zerweck, J., Knaute, T., Bräunlein, E., Samaras, P., Lautenbacher, L., Klaeger,
421 S., Wenschuh, H., Rad, R., Delanghe, B., Huhmer, A., Carr, S. A., Clauser, K. R., Krackhardt, A. M., Reimer,
422 U., and Kuster, B. (2021) Deep learning boosts sensitivity of mass spectrometry-based
423 immunopeptidomics. *Nat. Commun. 2021 121* 12, 1–12

424 25. Bouwmeester, R., Gabriels, R., Hulstaert, N., Martens, L., and Degroeve, S. (2021) DeepLC can predict
425 retention times for peptides that carry as-yet unseen modifications. *Nat. Methods 2021*, 1–7

426 26. Gabriels, R., Martens, L., and Degroeve, S. (2019) Updated MS^2PIP web server delivers fast and accurate
427 MS^2 peak intensity prediction for multiple fragmentation methods, instruments and labeling techniques.
428 *Nucleic Acids Res.* 47, W295–W299

429 27. Degroeve, S., Maddelein, D., and Martens, L. (2015) MS2PIP prediction server: compute and visualize
430 MS2 peak intensity predictions for CID and HCD fragmentation. *Nucleic Acids Res.* 43, W326-30

431 28. Degroeve, S., Martens, L., and Jurisica, I. (2013) MS2PIP: A tool for MS/MS peak intensity prediction.
432 *Bioinformatics* 29, 3199–3203

433 29. Ruiz Cuevas, M. V., Hardy, M.-P., Holly, J., Bonneil, É., Durette, C., Courcelles, M., Lanoix, J., Côté, C.,
434 Staudt, L. M., Lemieux, S., Thibault, P., Perreault, C., and Yewdell, J. (2020) Most Non-Canonical Proteins
435 Uniquely Populate the Proteome or Immunopeptidome. *SSRN Electron. J.*,

436 30. Martens, L., Hermjakob, H., Jones, P., Adamsk, M., Taylor, C., States, D., Gevaert, K., Vandekerckhove, J.,
437 and Apweiler, R. (2005) PRIDE: The proteomics identifications database. *Proteomics*,

438 31. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D. J., Inuganti, A.,
439 Griss, J., Mayer, G., Eisenacher, M., Pérez, E., Uszkoreit, J., Pfeuffer, J., Sachsenberg, T., Yılmaz, Ş., Tiwary,
440 S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A. F., Ternent, T., Brazma, A., and Vizcaíno, J. A. (2019) The
441 PRIDE database and related tools and resources in 2019: improving support for quantification data.
442 *Nucleic Acids Res.* 47, D442–D450

443 32. Hulstaert, N., Shofstahl, J., Sachsenberg, T., Walzer, M., Barsnes, H., Martens, L., and Perez-Riverol, Y.
444 (2020) ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *J. Proteome*
445 *Res.* 19, 537–542

446 33. Deutsch, E. W., Perez-Riverol, Y., Carver, J., Kawano, S., Mendoza, L., Den Bossche, T. Van, Gabriels, R.,
447 Binz, P. A., Pullman, B., Sun, Z., Shofstahl, J., Bittremieux, W., Mak, T. D., Klein, J., Zhu, Y., Lam, H.,
448 Vizcaíno, J. A., and Bandeira, N. (2020) Universal spectrum identifier for mass spectra. *bioRxiv*,
449 2020.12.07.415539

450 34. Chen, T., and Guestrin, C. *XGBoost: A Scalable Tree Boosting System* (ACM, New York, NY, USA)

451 35. Bergstra, J., Yamins, D., and Cox, D. D. (2013) Making a Science of Model Search: Hyperparameter
452 Optimization in Hundreds of Dimensions for Vision Architectures. 28,

453 36. Granholm, V., Kim, S., Navarro, J. C. F., Sjölund, E., Smith, R. D., and Käll, L. (2014) Fast and accurate
454 database searches with MS-GF+percolator. *J. Proteome Res.* 13, 890–897

455 37. Tyanova, S., Temu, T., and Cox, J. (2016) The MaxQuant computational platform for mass spectrometry-
456 based shotgun proteomics. *Nat. Protoc. 2016 1112* 11, 2301–2319

457 38. Sarkizova, S., Klaeger, S., Le, P. M., Li, L. W., Oliveira, G., Keshishian, H., Hartigan, C. R., Zhang, W., Braun,
458 D. A., Ligon, K. L., Bachireddy, P., Zervantonakis, I. K., Rosenbluth, J. M., Ouspenskaia, T., Law, T., Justesen,
459 S., Stevens, J., Lane, W. J., Eisenhaure, T., Lan Zhang, G., Clauser, K. R., Hacohen, N., Carr, S. A., Wu, C. J.,
460 and Keskin, D. B. (2020) A large peptidome dataset improves HLA class I epitope prediction across most

461 of the human population. *Nat. Biotechnol.* 38, 199–209

462 39. Hunter, J. D. (2007) Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* 9, 90–95

463 40. Waskom, M. (2021) seaborn: statistical data visualization. *J. Open Source Softw.* 6, 3021

464 41. Bittremieux, W. (2019) spectrum_utils: A Python Package for Mass Spectrometry Data Processing and
465 Visualization. *Anal. Chem.* 92, 659–661

466 42. Chen, R., Fauteux, F., Foote, S., Stupak, J., Tremblay, T. L., Gurnani, K., Fulton, K. M., Weeratna, R. D.,
467 Twine, S. M., and Li, J. (2018) Chemical Derivatization Strategy for Extending the Identification of MHC
468 Class i Immunopeptides. *Anal. Chem.* 90, 11409–11416

469 43. Racle, J., Michaux, J., Rockinger, G. A., Arnaud, M., Bobisse, S., Chong, C., Guillaume, P., Coukos, G.,
470 Harari, A., Jandus, C., Bassani-Sternberg, M., and Gfeller, D. (2019) Robust prediction of HLA class II
471 epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* 37, 1283–1286

472 44. Chong, C., Marino, F., Pak, H., Racle, J., Daniel, R. T., Mü Ller, M., Gfeller, D., Coukos, G., and Bassani-
473 Sternberg, M. (2018) High-throughput and sensitive immunopeptidomics platform reveals profound
474 interferon γ-mediated remodeling of the human leukocyte antigen (HLA) ligandome. *Mol. Cell.*
475 *Proteomics* 17, 533–548

476 45. Gfeller, D., Guillaume, P., Michaux, J., Pak, H.-S., Daniel, R. T., Racle, J., Coukos, G., and Bassani-Sternberg,
477 M. (2018) The Length Distribution and Multiple Specificity of Naturally Presented HLA-I Ligands. *J.*
478 *Immunol.* 201, 3705–3716

479 46. Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber,
480 J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., Busch, D. H., Peschel, C., Rad,
481 R., Cox, J., Mann, M., Krackhardt, A. M., M, B.-S., E, B., R, K., T, E., P, S., S, A., M, S., J, W., J, S.-H., K, S.,
482 ME, M., A, W., R, H., D, H. B., C, P., R, R., J, C., M, M., and AM, K. (2016) Direct identification of clinically
483 relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat.*
484 *Commun.* 7, 1–16

485 47. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., Zecha, J., Asplund, A., Li, L., Meng,
486 C., Frejno, M., Schmidt, T., Schnatbaum, K., Wilhelm, M., Ponten, F., Uhlen, M., Gagneur, J., Hahne, H.,
487 and Kuster, B. (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues.
488 *Mol. Syst. Biol.* 15,

489 48. Bassani-Sternberg, M., Chong, C., Guillaume, P., Solleder, M., Pak, H., Gannon, P. O., Kandalaft, L. E.,
490 Coukos, G., and Gfeller, D. (2017) Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen
491 predictions and identifies allostery regulating HLA specificity. *bioRxiv* 13, 098780

492 49. Marino, F., Semilietof, A., Michaux, J., Pak, H. S., Coukos, G., Müller, M., and Bassani-Sternberg, M. (2020)
493 Biogenesis of HLA Ligand Presentation in Immune Cells Upon Activation Reveals Changes in Peptide
494 Length Preference. *Front. Immunol.* 11,

495 50. Gravina, F., Sanchuki, H. S., Rodrigues, T. E., Gerhardt, E. C. M., Pedrosa, F. O., Souza, E. M., Valdameri,
496 G., de Souza, G. A., and Huergo, L. F. (2018) Proteome analysis of an Escherichia coli ptsN-null strain
497 under different nitrogen regimes. *J. Proteomics* 174, 28–35

498

Supplementary information to

MS²Rescore: Data-driven rescoring dramatically boosts immunopeptide identification rates

Arthur Declercq[1,2], Robbin Bouwmeester[1,2], Sven Degroeve[1,2], Lennart Martens[1,2 §], and Ralf Gabriels[1,2]

[1] VIB-UGent Center for Medical Biotechnology, VIB, Belgium
[2] Department of Biomolecular Medicine, Ghent University, Belgium

[§] To whom correspondence should be addressed
Tel: +32 9 264 93 58
Email: lennart.martens@vib-ugent.be
Address: Technologiepark 75, 9052 Ghent, Belgium

ORCID IDs:
Arthur Declercq: https://orcid.org/0000-0002-9376-1399/
Robbin Bouwmeester: https://orcid.org/0000-0001-6807-7029/
Sven Degroeve: https://orcid.org/0000-0001-8349-3370/
Lennart Martens: https://orcid.org/0000-0003-4277-658X/
Ralf Gabriels: https://orcid.org/0000-0002-1679-1711/

527 Supplementary Table S1: List of all PRIDE Archive projects used for training, testing, and evaluating new
528 MS²PIP models. Unique peptides denote the total number of unique peptides in terms of sequence,
529 modifications, and precursor charge within each project.

PRIDE Archive project	Unique peptides	Peptide type	Publication
		Train / test	
PXD012308	67 922	Immunopeptide (HLA-II)	J. Racle et al.(43)
PXD006939	229 265	Immunopeptide (HLA-I/II)	C. Chong et al.(44)
PXD009925	40 920	Immunopeptide (HLA-I)	D. Gfeller et al. (45)
PXD000394	149 677	Immunopeptide (HLA-I)	M. Bassani-Sternberg et al. (7)
PXD004894	129 011	Immunopeptide (HLA-I/II)	M. Bassani-Sternberg et al. (46)
PXD010154 (70%)	60 388	Chymotrypsin-digested	D. Wang et al. (47)
		Evaluation	
PXD005231	46 797	Immunopeptide (HLA-I)	M. Bassani-Sternberg et al. (48)
PXD020011	23 982	Immunopeptide (HLA-II)	F. Marino et al. (49)
PXD010154 (30%)	25 980	Chymotrypsin-digested	D. Wang et al. (47)
PXD008034	35 269	General proteomics data	F. Gravina et al. (50)

530

Supplementary Table S2: The optimal hyperparameters for each new b- and y-ion MS²PIP model, as determined during hyperparameter optimization.

Model	Eta	Max depth	Grow policy	Max leaves	Min child weight	Gamma	Lambda	Alpha	Colsample by tree	Sub-sample
Immuno-chymotrypsin (b-ions)	0.08060612330262913	18	Lossguide	117	500	0.031142279181653326	0.2724553826622634	3.4	0.891381182690278	0.7
Immuno-chymotrypsin (y-ions)	0.047107785048838	18	Lossguide	490	4	0.37528441949267444	0.35150807248415	3.3	0.6122042447952851	0.6
Immunopeptide (b-ions)	0.09263630381479264	17	Lossguide	131	16	0.6048882172751935	0.9332236183206803	4.6	0.9898165069470042	0.7
Immunopeptide (y-ions)	0.0594145790364741	17	Lossguide	302	3	0.03338151150211477	0.4430375595950531	4.5	0.9389820388602939	0.7
Non-tryptic immunopeptide (b-ions)	0.05754126608420542	16	Lossguide	356	85	0.35987834628709725	0.7049467859206121	3.5	0.7002269364273864	0.6
Non-tryptic immunopeptide (y-ions)	0.08443065926710414	18	Lossguide	6	7	0.38091488564428677	0.6137299313065955	4.3	0.7958195937743725	1.0

Supplementary Table S3: Percolator rescoring features generated by MS²Rescore from MaxQuant search results.

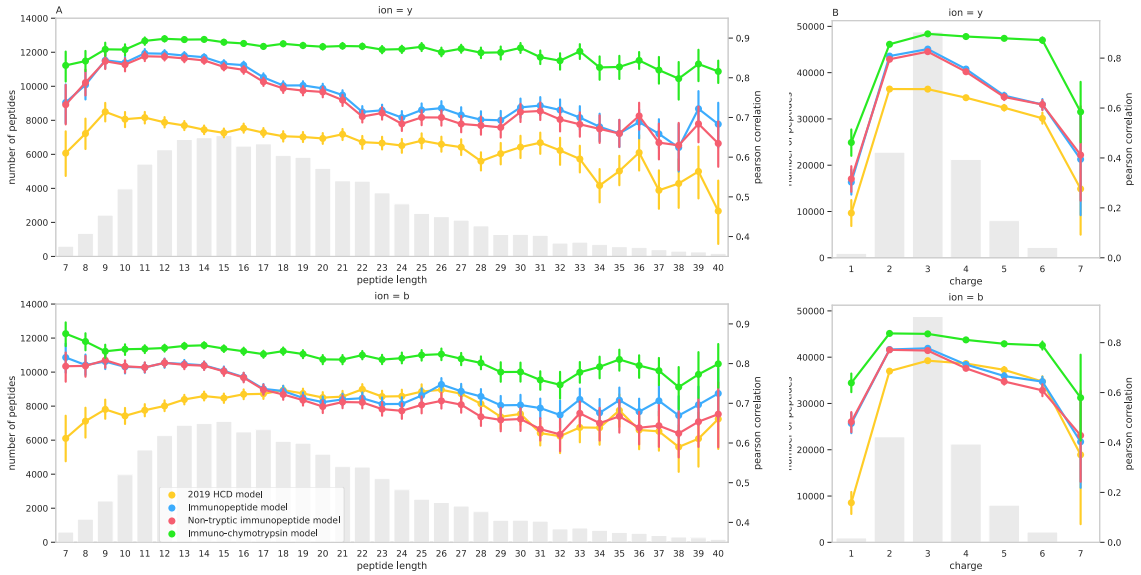
Feature name	Description
	Search engine features (MaxQuant integration)
ln_ms2_ion_current	Natural logarithm of the summed intensity of all annotated fragment ions
ln_cterm_ion_current_ratio	Natural logarithm of the summed intensity of the annotated y-ions, divided by that of all annotated fragment ions
ln_nterm_ion_current_ratio	Natural logarithm of the summed intensity of the annotated b-ions, divided by that of all annotated fragment ions
ln_explained_ion_current	Natural logarithm of the summed intensity of the annotated fragment ions, divided by that of all fragment ions
stdev_error_top7	Standard deviation of the mass errors of the seven most intense annotated fragment ions
sq_mean_error_top7	Squared mean of the mass errors of the seven most intense annotated fragment ions
mean_error_top7	Mean of the mass errors of the seven most intense annotated fragment ions
Charge7	One-hot encoded precursor charge state (1 if charge is 7, else 0)
Charge6	One-hot encoded precursor charge state (1 if charge is 6, else 0)
Charge5	One-hot encoded precursor charge state (1 if charge is 5, else 0)
Charge4	One-hot encoded precursor charge state (1 if charge is 4, else 0)
Charge3	One-hot encoded precursor charge state (1 if charge is 3, else 0)
Charge2	One-hot encoded precursor charge state (1 if charge is 2, else 0)
Charge1	One-hot encoded precursor charge state (1 if charge is 1, else 0)
absdM	Absolute difference between theoretical and observed precursor mass
enzInt	Number of missed enzymatic cleavages
dM	Difference between theoretical and observed precursor mass
PepLen	Peptide length
Mass	Precursor mass
ChargeN	Precursor charge state
RawModLocProb	Andromeda modification localization probability
RawDeltaScore	Difference in Andromeda score with next lower-ranking PSM
RawScore	Andromeda score
	DeepLC retention time features
predicted_retention_time_best	Predicted retention time for the precursor peak (all PSMs with the same sequence, modifications, and charge) closest to the predicted retention time
observed_retention_time_best	Observed retention time for the precursor peak (all PSMs with the same sequence, modifications, and charge) closest to the predicted retention time
rt_diff_best	Difference between observed and predicted retention time for the precursor peak (all PSMs with the same sequence, modifications, and charge) closest to the predicted retention time
rt_diff	Difference between observed and predicted retention time
predicted_retention_time	Predicted retention time
observed_retention_time	Observed retention time
	MS²PIP spectrum prediction features
cos_iony	Cosine similarity on y-ions
cos_ionb	Cosine similarity on b-ions
cos	Cosine similarity on b- and y-ions
dotprod_iony	Dot product on y-ions

dotprod_ionb	Dot product on b-ions
dotprod	Dot product on b- and y-ions
iony_std_abs_diff	Standard deviation of the absolute differences (y-ions only)
iony_mean_abs_diff	Mean of the absolute differences (y-ions only)
iony_abs_diff_Q3	Quantile 3 of the absolute differences (y-ions only)
iony_abs_diff_Q2	Quantile 2 of the absolute differences (y-ions only)
iony_abs_diff_Q1	Quantile 1 of the absolute differences (y-ions only)
iony_max_abs_diff	Maximum absolute difference (y-ions only)
iony_min_abs_diff	Minimum absolute difference (y-ions only)
ionb_std_abs_diff	Standard deviation of the absolute differences (b-ions only)
ionb_mean_abs_diff	Mean of the absolute differences (b-ions only)
ionb_abs_diff_Q3	Quantile 3 of the absolute differences (b-ions only)
ionb_abs_diff_Q2	Quantile 2 of the absolute differences (b-ions only)
ionb_abs_diff_Q1	Quantile 1 of the absolute differences (b-ions only)
ionb_max_abs_diff	Maximum absolute difference (b-ions only)
ionb_min_abs_diff	Minimum absolute difference (b-ions only)
std_abs_diff	Standard deviation of the absolute differences
mean_abs_diff	Mean of the absolute differences
abs_diff_Q3	Quantile 3 of the absolute differences
abs_diff_Q2	Quantile 2 of the absolute differences
abs_diff_Q1	Quantile 1 of the absolute differences
max_abs_diff	Maximum absolute difference
min_abs_diff	Minimum absolute difference
max_abs_diff_iontype	Ion type with maximum absolute difference
min_abs_diff_iontype	Ion type with minimum absolute difference
iony_mse	Mean square error on y-ions
ionb_mse	Mean square error on b-ions
spec_mse	Mean square error on b- and y-ions
iony_spearman	Spearman correlation on b- and y-ions
ionb_spearman	Spearman correlation on b- and y-ions
spec_spearman	Spearman correlation on b- and y-ions
iony_pearson	Pearson correlation coefficient on y-ions
ionb_pearson	Pearson correlation coefficient on b-ions
spec_pearson	Pearson correlation coefficient on b- and y-ions
cos_iony_norm	Cosine similarity on y-ions (log2-normalized spectrum)
cos_ionb_norm	Cosine similarity on b-ions (log2-normalized spectrum)
cos_norm	Cosine similarity on b- and y-ions (log2-normalized spectrum)
dotprod_iony_norm	Dot product on y-ions (log2-normalized spectrum)
dotprod_ionb_norm	Dot product on b-ions (log2-normalized spectrum)
dotprod_norm	Dot product on b- and y-ions (log2-normalized spectrum)
iony_std_abs_diff_norm	Standard deviation of the absolute differences (y-ions only) (log2-normalized spectrum)
iony_mean_abs_diff_norm	Mean of the absolute differences (y-ions only) (log2-normalized spectrum)
iony_abs_diff_Q3_norm	Quantile 3 of the absolute differences (y-ions only) (log2-normalized spectrum)
iony_abs_diff_Q2_norm	Quantile 2 of the absolute differences (y-ions only) (log2-normalized spectrum)
iony_abs_diff_Q1_norm	Quantile 1 of the absolute differences (y-ions only) (log2-normalized spectrum)

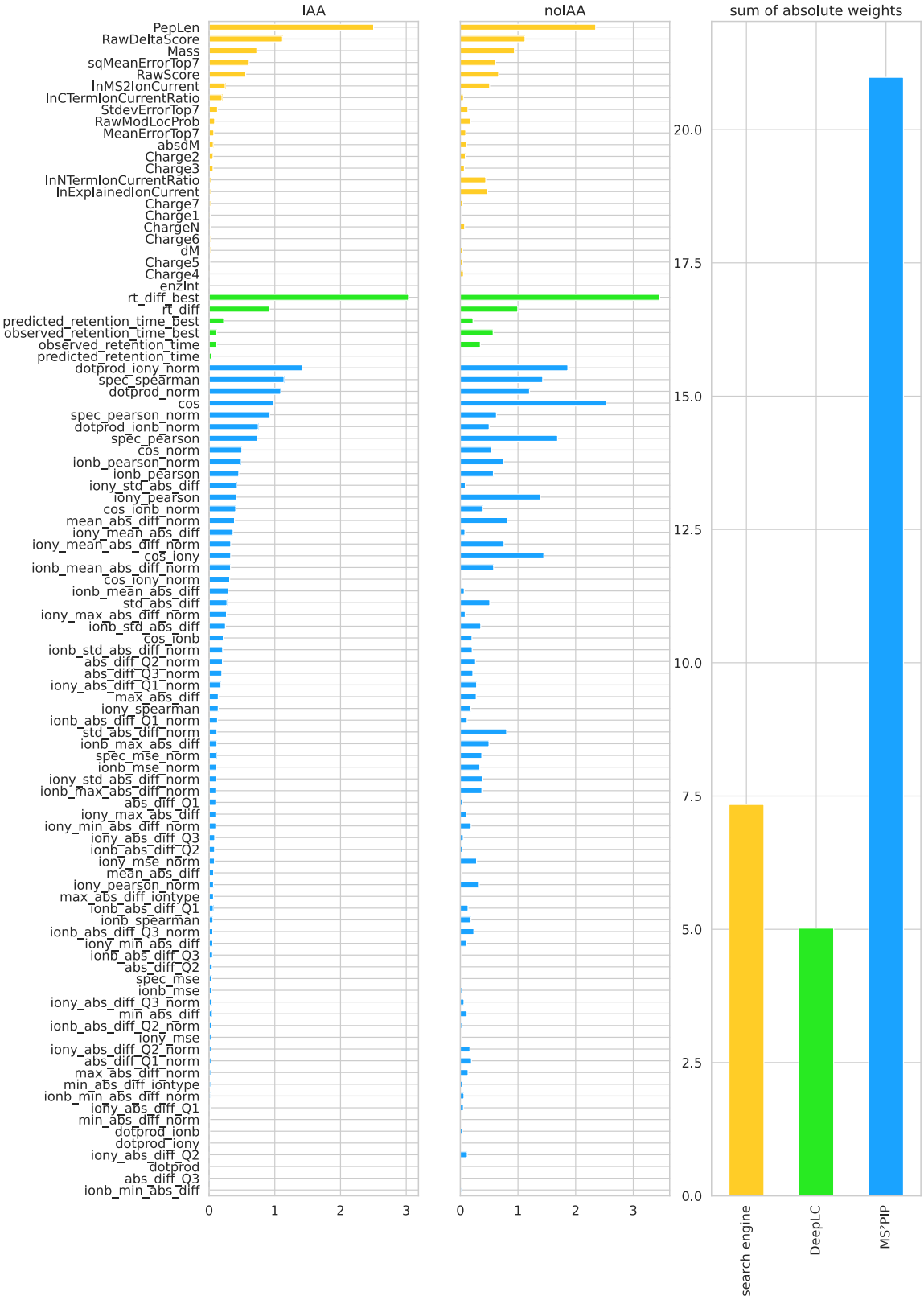
iony_max_abs_diff_norm	Maximum absolute difference (y-ions only) (log2-normalized spectrum)
iony_min_abs_diff_norm	Minimum absolute difference (y-ions only) (log2-normalized spectrum)
ionb_std_abs_diff_norm	Standard deviation of the absolute differences (b-ions only) (log2-normalized spectrum)
ionb_mean_abs_diff_norm	Mean of the absolute differences (b-ions only) (log2-normalized spectrum)
ionb_abs_diff_Q3_norm	Quantile 3 of the absolute differences (b-ions only) (log2-normalized spectrum)
ionb_abs_diff_Q2_norm	Quantile 2 of the absolute differences (b-ions only) (log2-normalized spectrum)
ionb_abs_diff_Q1_norm	Quantile 1 of the absolute differences (b-ions only) (log2-normalized spectrum)
ionb_max_abs_diff_norm	Maximum absolute difference (b-ions only) (log2-normalized spectrum)
ionb_min_abs_diff_norm	Minimum absolute difference (b-ions only) (log2-normalized spectrum)
std_abs_diff_norm	Standard deviation of the absolute differences (log2-normalized spectrum)
mean_abs_diff_norm	Mean of the absolute differences (log2-normalized spectrum)
abs_diff_Q3_norm	Quantile 3 of the absolute differences (log2-normalized spectrum)
abs_diff_Q2_norm	Quantile 2 of the absolute differences (log2-normalized spectrum)
abs_diff_Q1_norm	Quantile 1 of the absolute differences (log2-normalized spectrum)
max_abs_diff_norm	Ion type with maximum absolute difference (log2-normalized spectrum)
min_abs_diff_norm	Ion type with minimum absolute difference (log2-normalized spectrum)
iony_mse_norm	Mean square error on y-ions (log2-normalized spectrum)
ionb_mse_norm	Mean square error on b-ions (log2-normalized spectrum)
spec_mse_norm	Mean square error on b- and y-ions (log2-normalized spectrum)
iony_pearson_norm	Pearson correlation coefficient on y-ions (log2-normalized spectrum)
ionb_pearson_norm	Pearson correlation coefficient on b-ions (log2-normalized spectrum)
spec_pearson_norm	Pearson correlation coefficient on b- and y-ions (log2-normalized spectrum)

Supplementary Table S4: Median Pearson correlations for each model, ion-type, and evaluation data set.

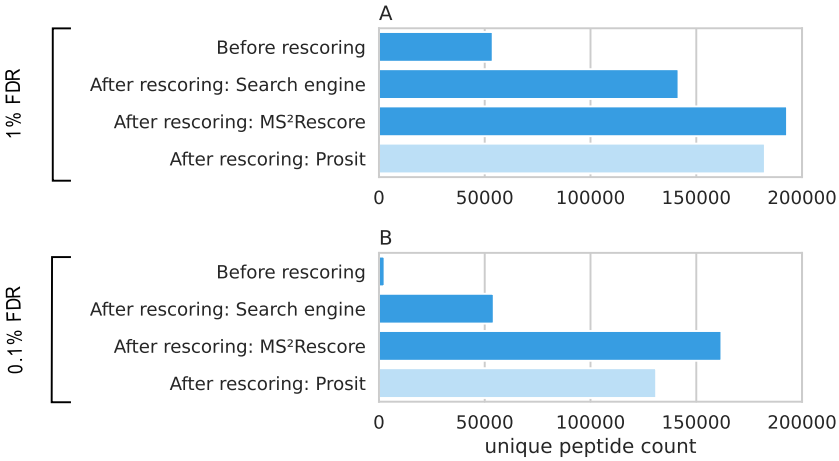
Model	Ion type	Evaluation data set	Median Pearson correlation
2019 HCD model	b	Chymotrypsin	0.789244
2019 HCD model	b	Immunopeptide	0.849500
2019 HCD model	b	General proteomics	0.905815
2019 HCD model	y	Chymotrypsin	0.727764
2019 HCD model	y	Immunopeptide	0.832574
2019 HCD model	y	General proteomics	0.901921
Immuno-chymotrypsin model	b	Chymotrypsin	0.886655
Immuno-chymotrypsin model	b	Immunopeptide	0.951049
Immuno-chymotrypsin model	b	General proteomics	0.926803
Immuno-chymotrypsin model	y	Chymotrypsin	0.925695
Immuno-chymotrypsin model	y	Immunopeptide	0.966121
Immuno-chymotrypsin model	y	General proteomics	0.937968
Immunopeptide model	b	Chymotrypsin	0.837299
Immunopeptide model	b	Immunopeptide	0.960748
Immunopeptide model	b	General proteomics	0.923750
Immunopeptide model	y	Chymotrypsin	0.869326
Immunopeptide model	y	Immunopeptide	0.966724
Immunopeptide model	y	General proteomics	0.936645
Non-tryptic model	b	Chymotrypsin	0.832734
Non-tryptic model	b	Immunopeptide	0.949395
Non-tryptic model	b	General proteomics	0.899387
Non-tryptic model	y	Chymotrypsin	0.861495
Non-tryptic model	y	Immunopeptide	0.954444
Non-tryptic model	y	General proteomics	0.864106



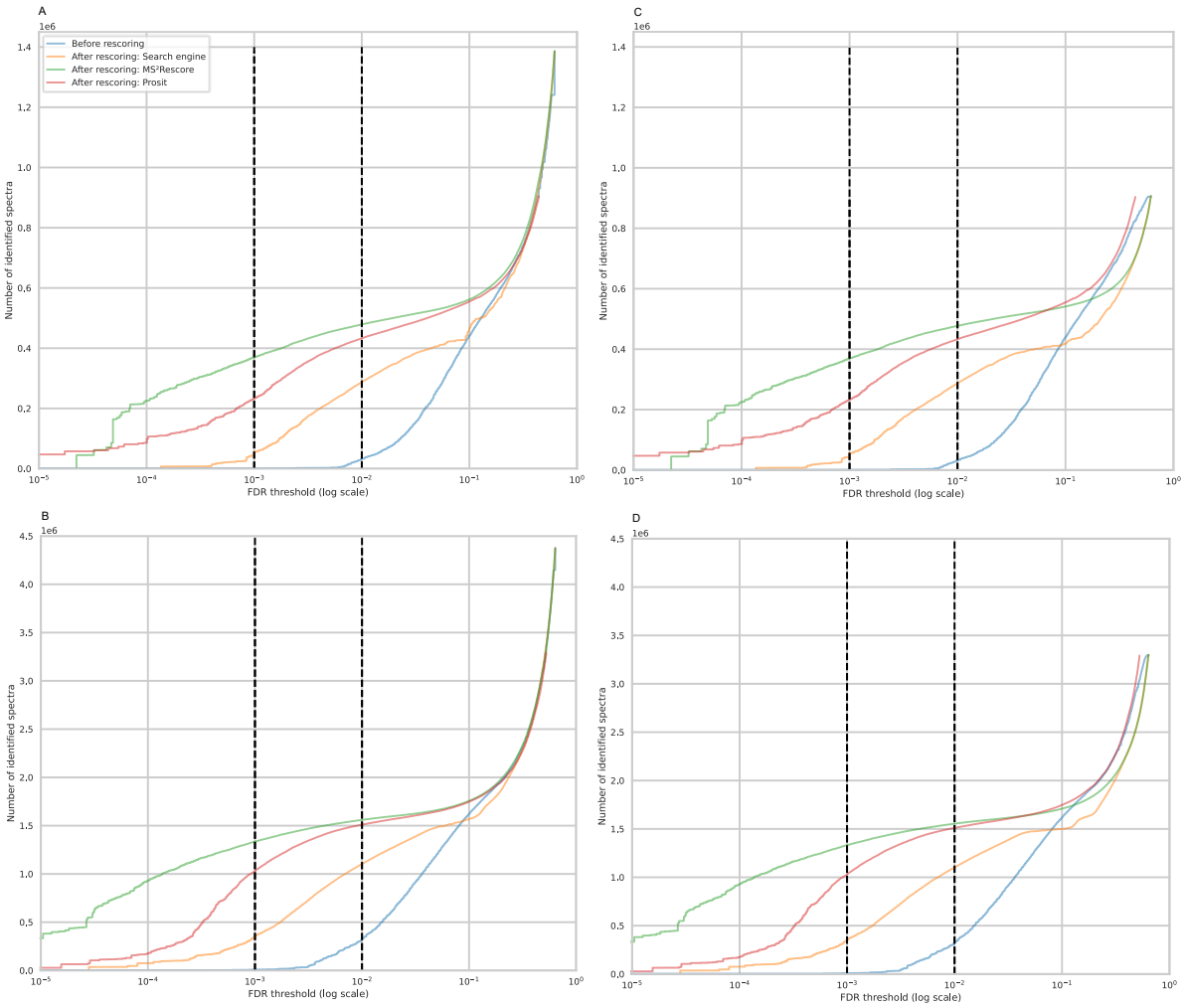
Supplementary Figure S1: MS²PIP model evaluation on chymotrypsin evaluation data by peptide length (A) and precursor charge (B). The connected dots indicate the median Pearson correlation for each MS²PIP y-ion model (top) and b-ion model (bottom), split by peptide length (left) or precursor charge (right). Error bars denote the Pearson correlation standard deviation within each group. The grey bar charts indicate the number of evaluation peptides within each group.



Supplementary Figure S2: Bar charts showing the mean Percolator feature weights (across the three Percolator cross validation splits) for each feature as generated by MS²Rescore, split by alkylated samples (IAA) on the left and non-alkylated samples (noIAA) in the middle. Blue bars denote MS²PIP based-features, green bars denote DeepLC-based features and yellow bars denote search engine derived features. The bar chart on the right shows the combined weight of each category, calculated by summing the respective absolute weights across both samples.



Supplementary Figure S3: Bar charts showing total number of unique identified peptides in terms of sequence for each rescoring method at 1% FDR (A) and at 0.1% FDR (B).



Supplementary Figure S4: Comparison of the total number of identified spectra for different rescoring methods at varying FDR thresholds in log10 scale. Shown rescoring methods used are: Raw MaxQuant (Andromeda) scores without rescoring (*Before rescoring*, blue), rescoring with only search engine feature as Percolator would do (*After rescoring: Search engine*, orange), full MS²Rescore rescoring with search engine-, DeepLC-, and MS²PIP-features (*After rescoring: MS²Rescore*, green) and rescoring with Prosit as performed by the original authors (*After rescoring: Prosit*, red). The dotted black lines denote the 0.1% false discovery rate threshold (left) and 1% false discovery rate threshold (right). Figures A and C show results for the non-alkylated samples (noIAA), figures B and D results for the alkylated samples (IAA). Figures on the left show results including all identifications compatible with MS²Rescore, figures on the right show results filtered down to only including identifications compatible with Prosit, which is a subset of the former.

Supplementary references:

7. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L. J., and Mann, M. (2015) Mass spectrometry of human leukocyte antigen class i peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics* 14, 658–673

43. Racle, J., Michaux, J., Rockinger, G. A., Arnaud, M., Bobisse, S., Chong, C., Guillaume, P., Coukos, G., Harari, A., Jandus, C., Bassani-Sternberg, M., and Gfeller, D. (2019) Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* 37, 1283–1286

44. Chong, C., Marino, F., Pak, H., Racle, J., Daniel, R. T., Mü Ller, M., Gfeller, D., Coukos, G., and Bassani-Sternberg, M. (2018) High-throughput and sensitive immunopeptidomics platform reveals profound interferon γ-mediated remodeling of the human leukocyte antigen (HLA) ligandome. *Mol. Cell. Proteomics* 17, 533–548

45. Gfeller, D., Guillaume, P., Michaux, J., Pak, H.-S., Daniel, R. T., Racle, J., Coukos, G., and Bassani-Sternberg, M. (2018) The Length Distribution and Multiple Specificity of Naturally Presented HLA-I Ligands. *J. Immunol.* 201, 3705–3716

46. Bassani-Sternberg, M., Bräunlein, E., Klar, R., Engleitner, T., Sinitcyn, P., Audehm, S., Straub, M., Weber, J., Slotta-Huspenina, J., Specht, K., Martignoni, M. E., Werner, A., Hein, R., Busch, D. H., Peschel, C., Rad, R., Cox, J., Mann, M., Krackhardt, A. M., M, B.-S., E, B., R, K., T, E., P, S., S, A., M, S., J, W., J, S.-H., K, S., ME, M., A, W., R, H., D, H. B., C, P., R, R., J, C., M, M., and AM, K. (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* 7, 1–16

47. Wang, D., Eraslan, B., Wieland, T., Hallström, B., Hopf, T., Zolg, D. P., Zecha, J., Asplund, A., Li, L., Meng, C., Frejno, M., Schmidt, T., Schnatbaum, K., Wilhelm, M., Ponten, F., Uhlen, M., Gagneur, J., Hahne, H., and Kuster, B. (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.* 15,

48. Bassani-Sternberg, M., Chong, C., Guillaume, P., Solleder, M., Pak, H., Gannon, P. O., Kandalaft, L. E., Coukos, G., and Gfeller, D. (2017) Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *bioRxiv* 13, 098780

49. Marino, F., Semilietof, A., Michaux, J., Pak, H. S., Coukos, G., Müller, M., and Bassani-Sternberg, M. (2020) Biogenesis of HLA Ligand Presentation in Immune Cells Upon Activation Reveals Changes in Peptide Length Preference. *Front. Immunol.* 11,

50. Gravina, F., Sanchuki, H. S., Rodrigues, T. E., Gerhardt, E. C. M., Pedrosa, F. O., Souza, E. M., Valdameri, G., de Souza, G. A., and Huergo, L. F. (2018) Proteome analysis of an Escherichia coli ptsN-null strain under different nitrogen regimes. *J. Proteomics* 174, 28–35