

DEEP LEARNING TO UNCOVER THE IMMUNOPEPTIDOME

Wout Bittremieux — BMSS 2023



@wout@sigmoid.social

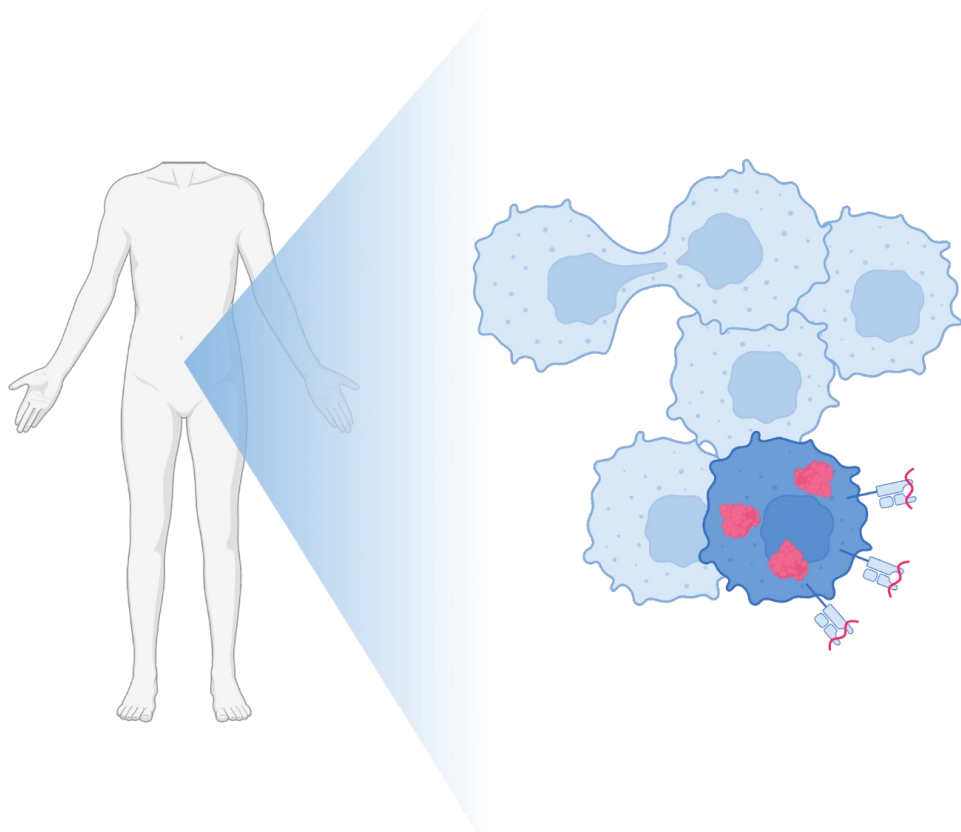


bittremieux.be

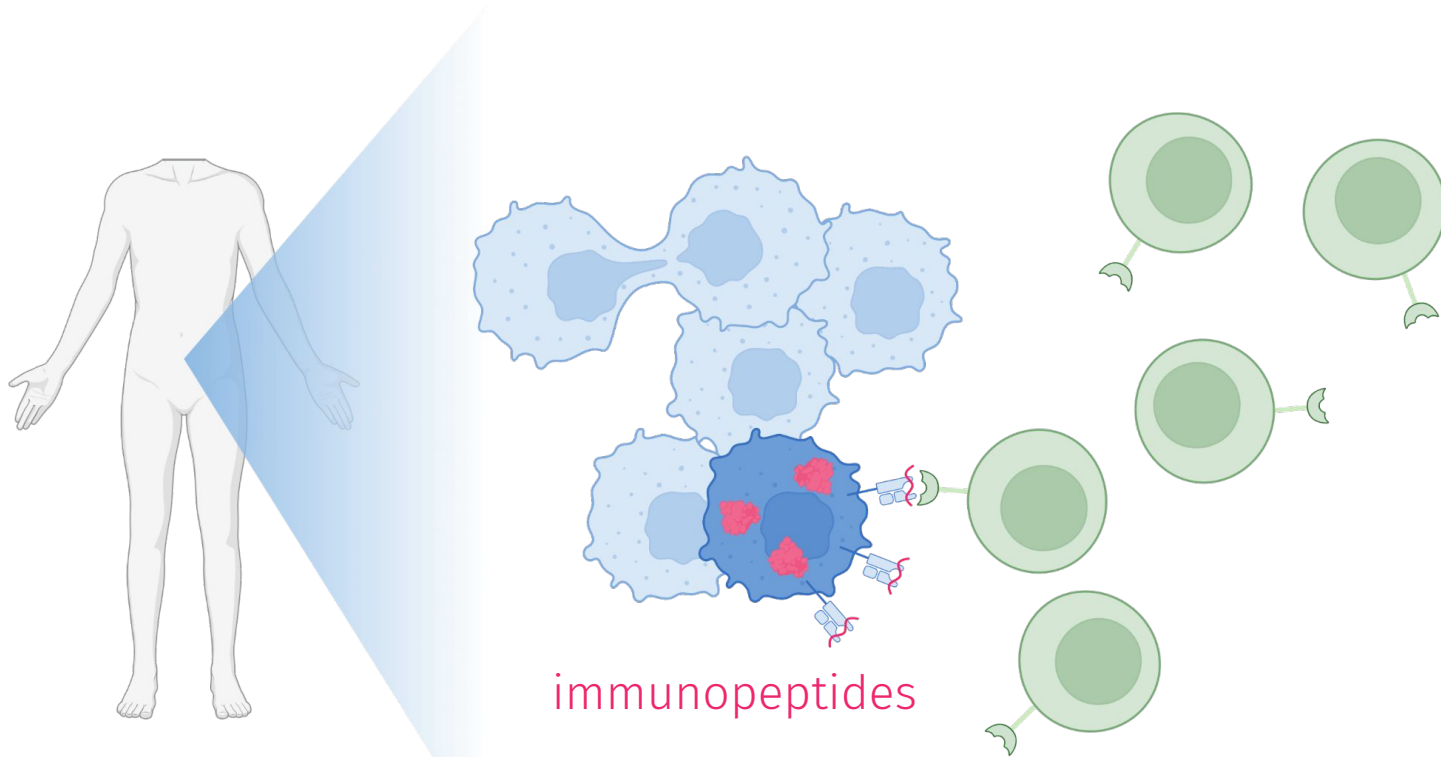


wout.bittremieux@uantwerpen.be

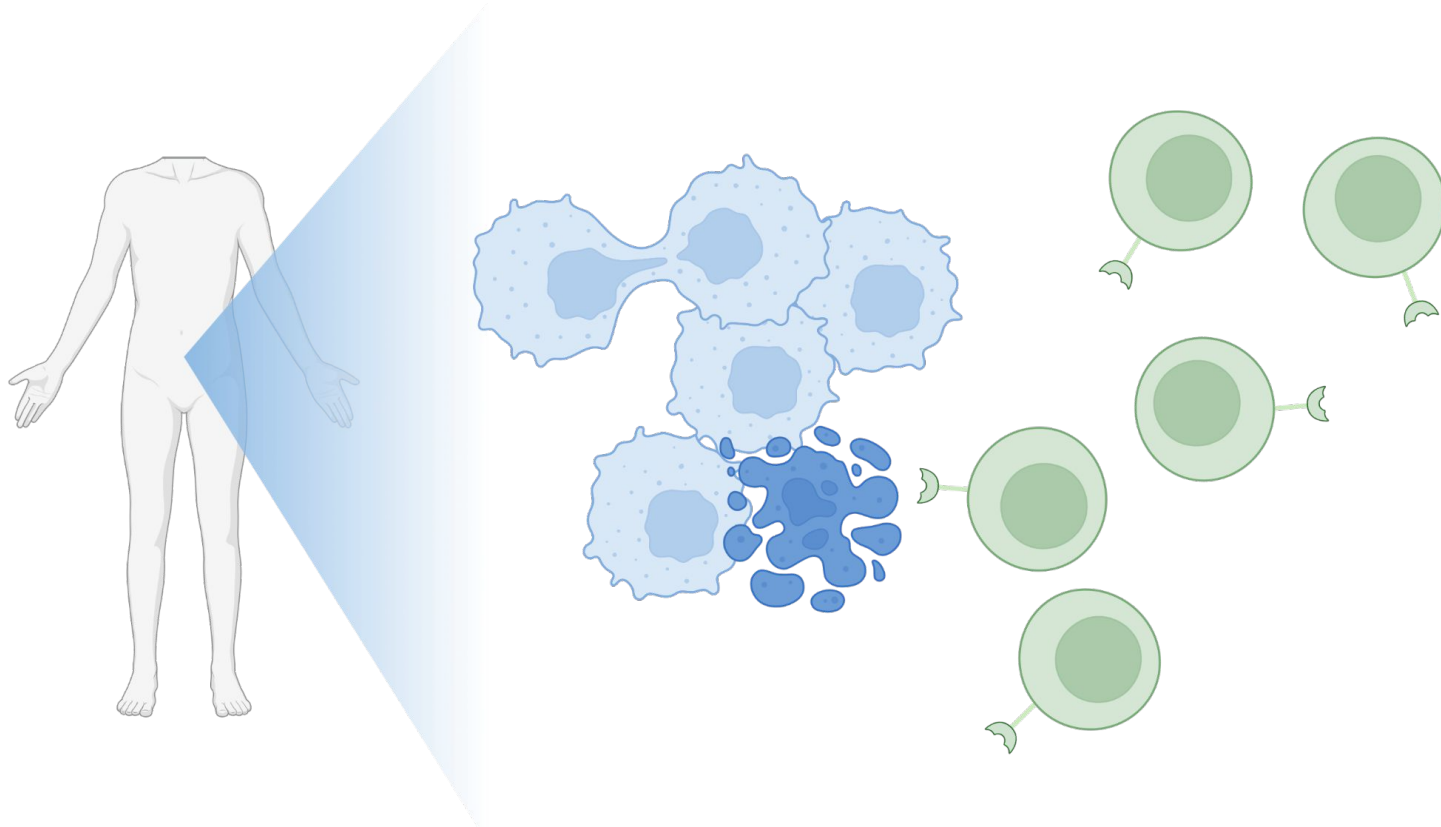
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



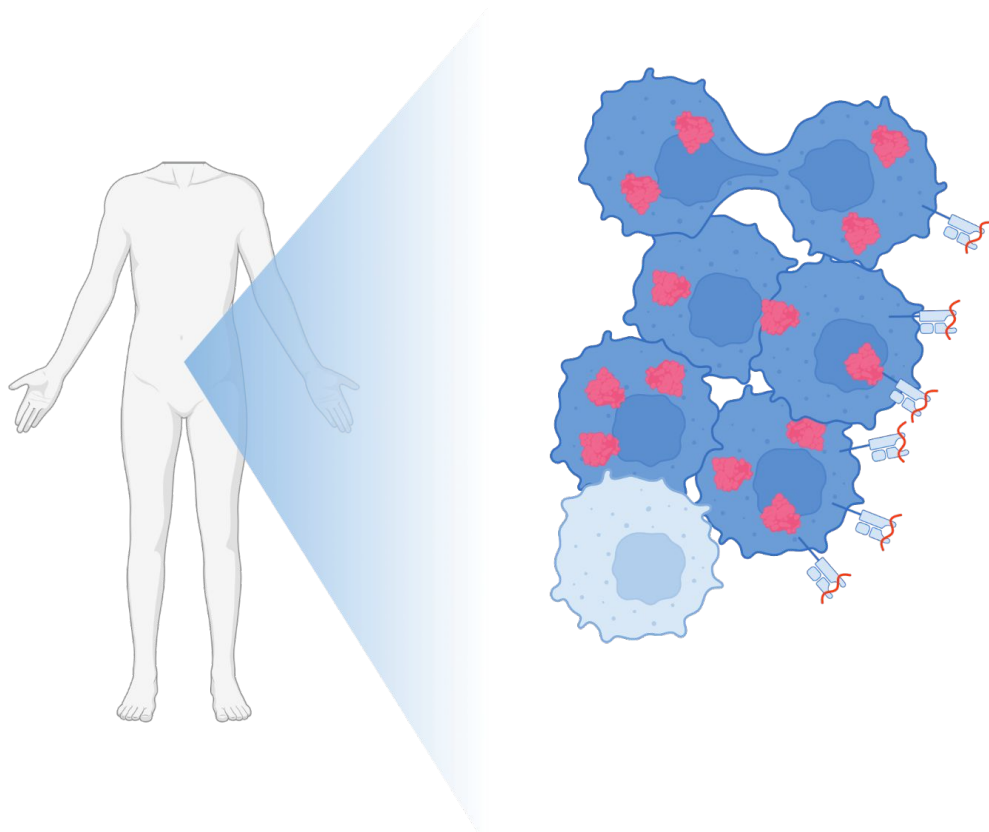
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



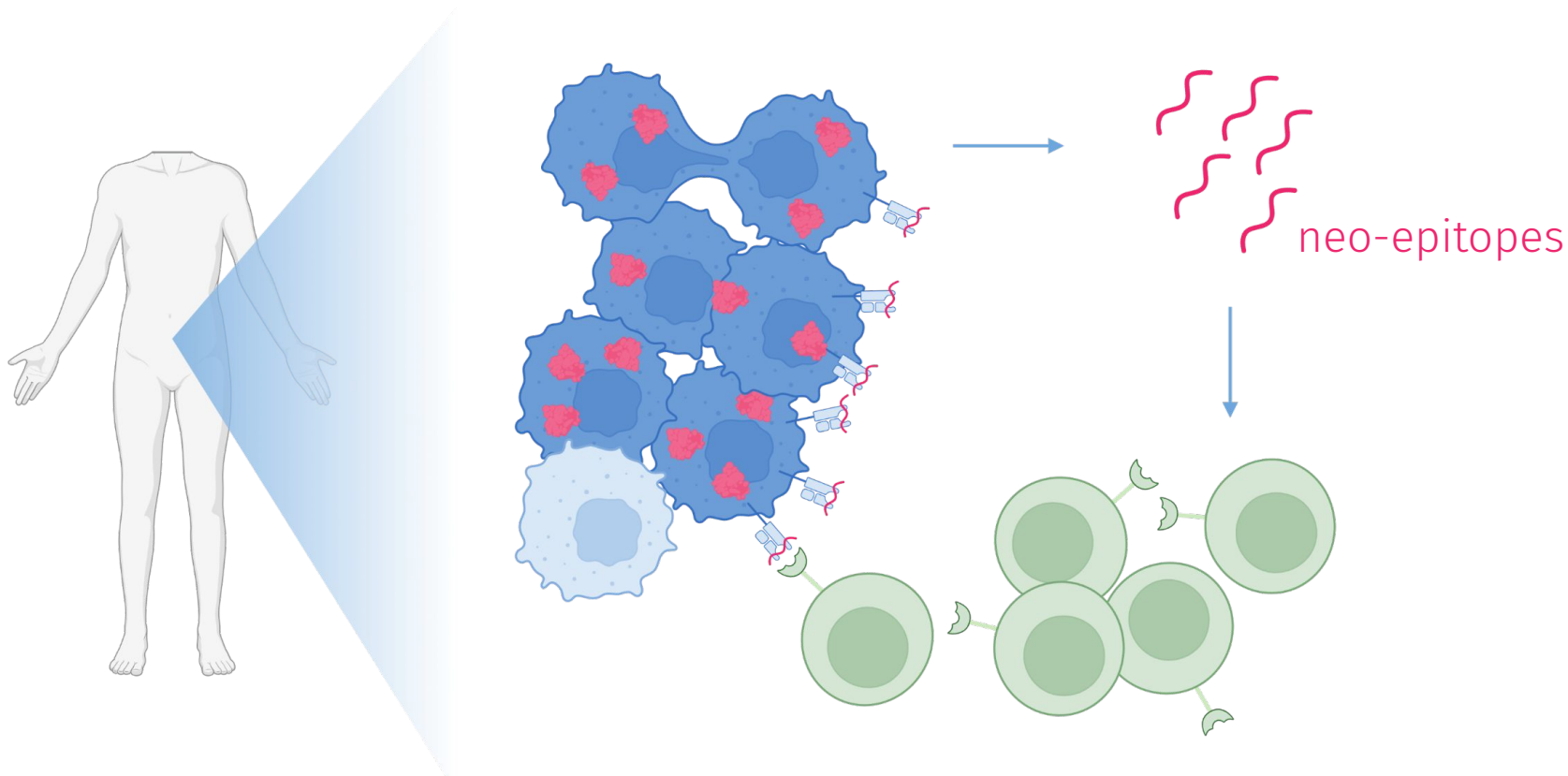
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



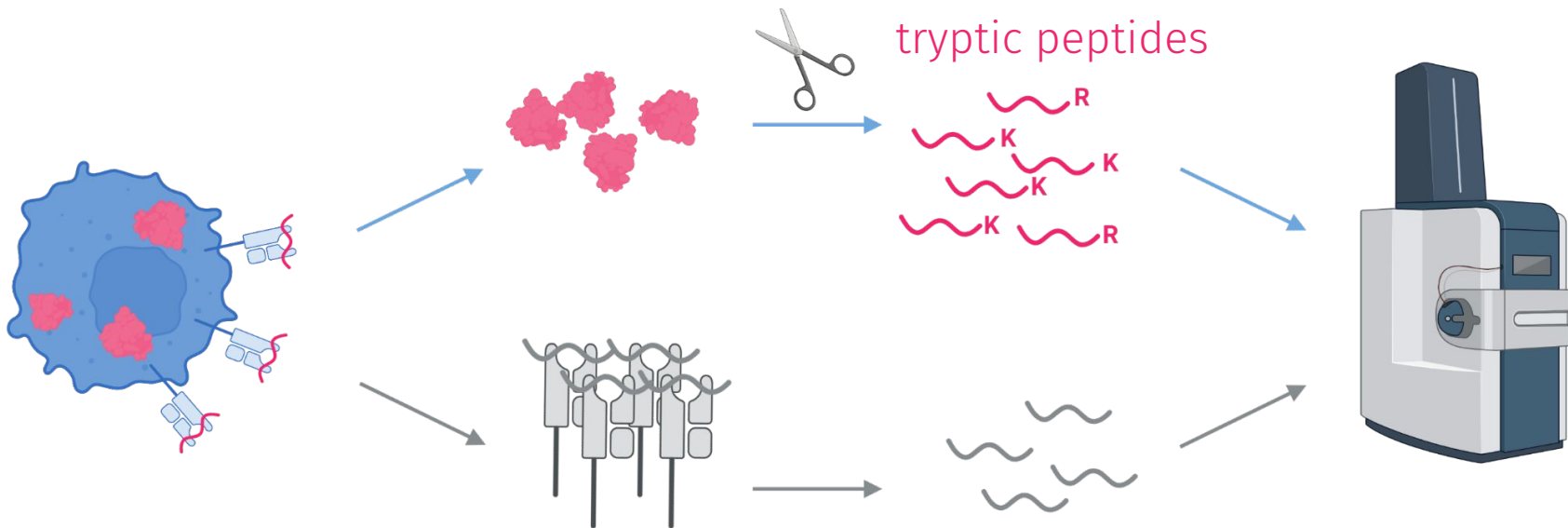
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



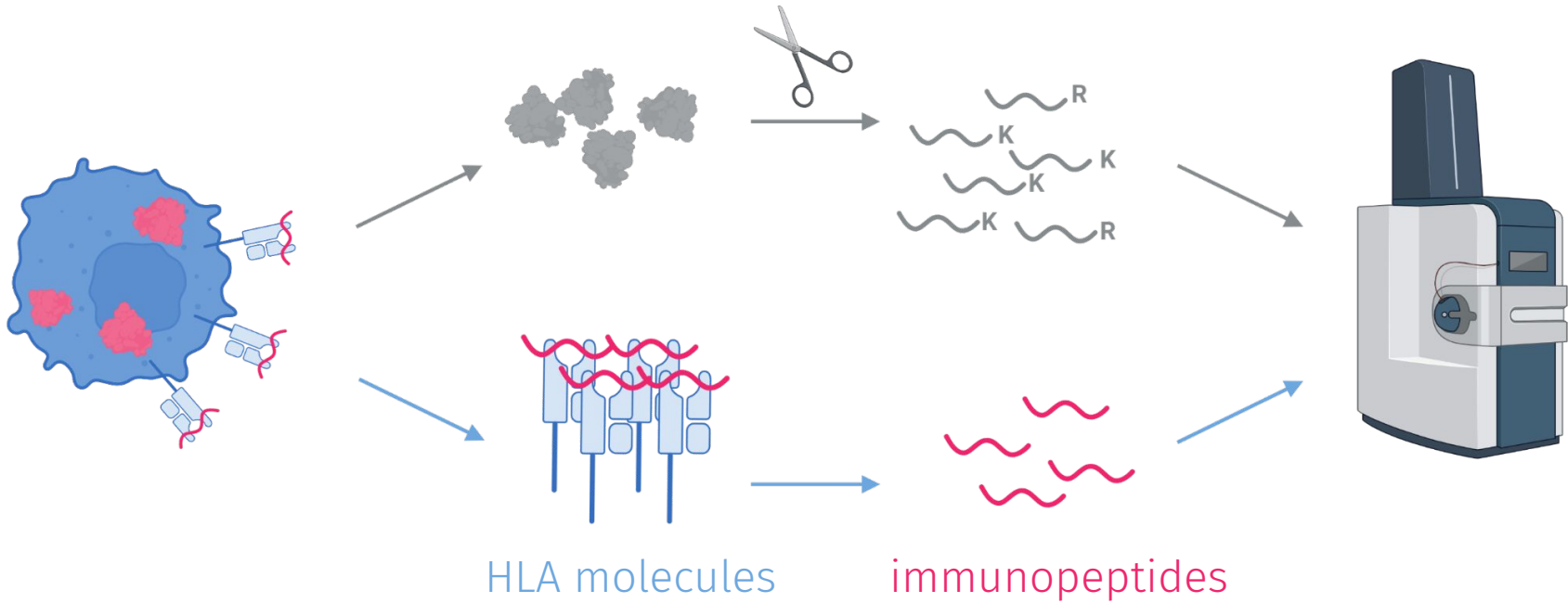
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



IMMUNOPEPTIDOMICS

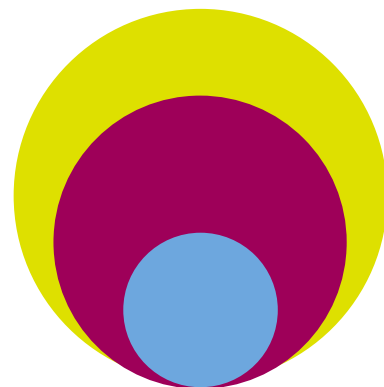


IMMUNOPEPTIDOMICS



SEARCH SPACE IN IMMUNOPEPTIDOMICS

- Massive search space: all protein subsequences have to be considered
- Increased probability of identifying high-scoring decoys
- Reduced identification rate at a fixed FDR



non-tryptic

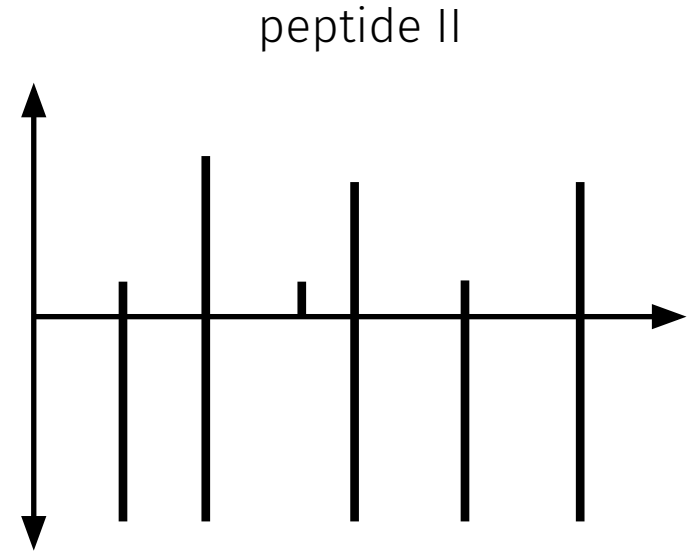
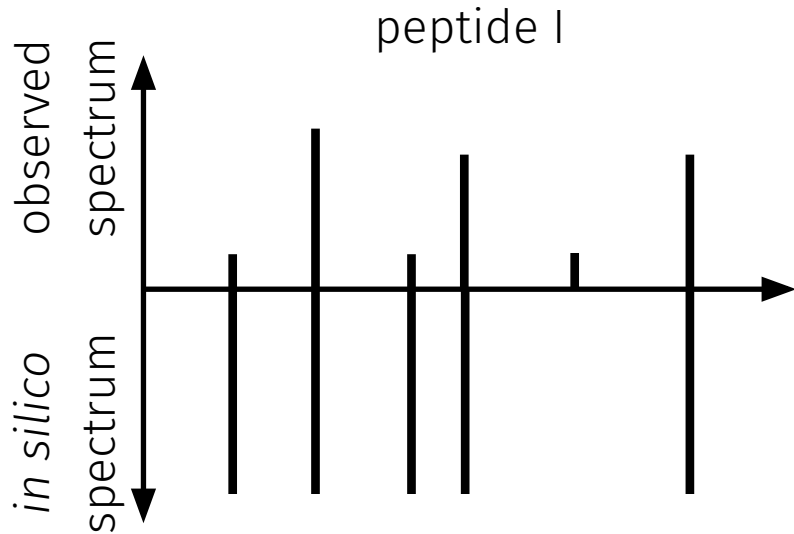
semi-tryptic

tryptic

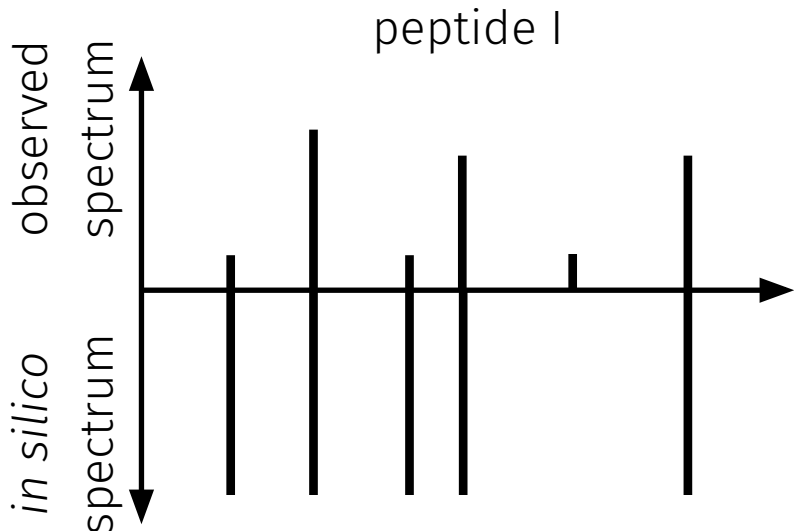


**SOLUTION 1: PREDICTING
FRAGMENT ION INTENSITIES**

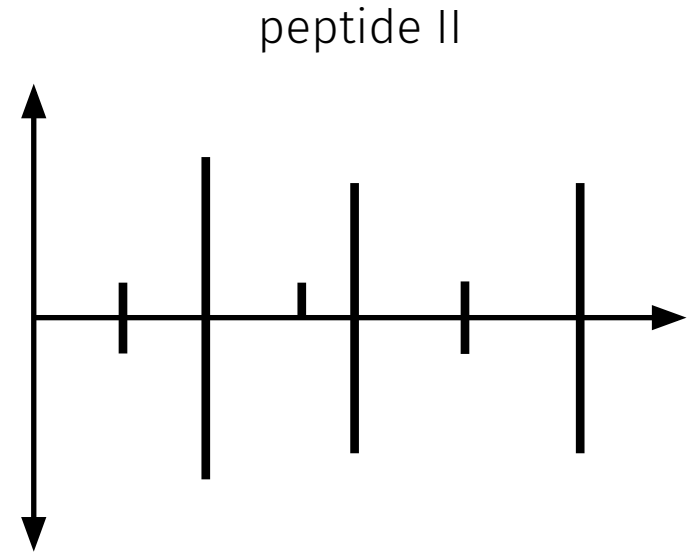
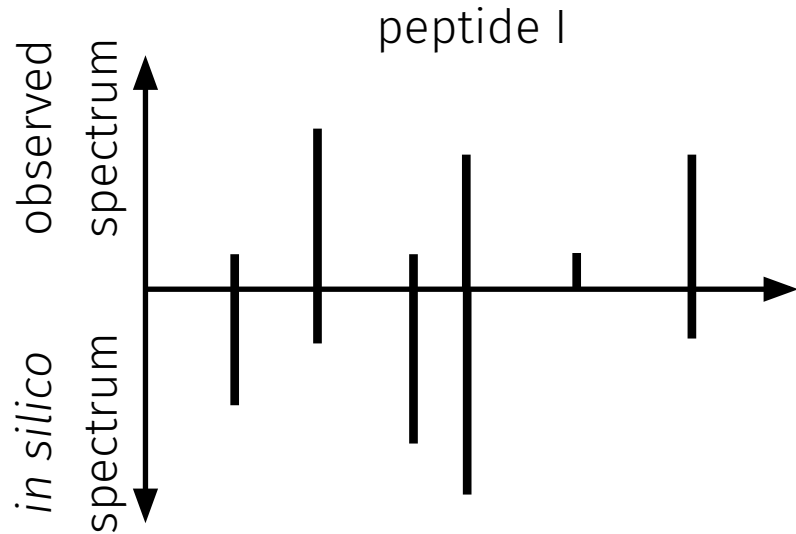
SPECTRUM ANNOTATION IS CHALLENGING



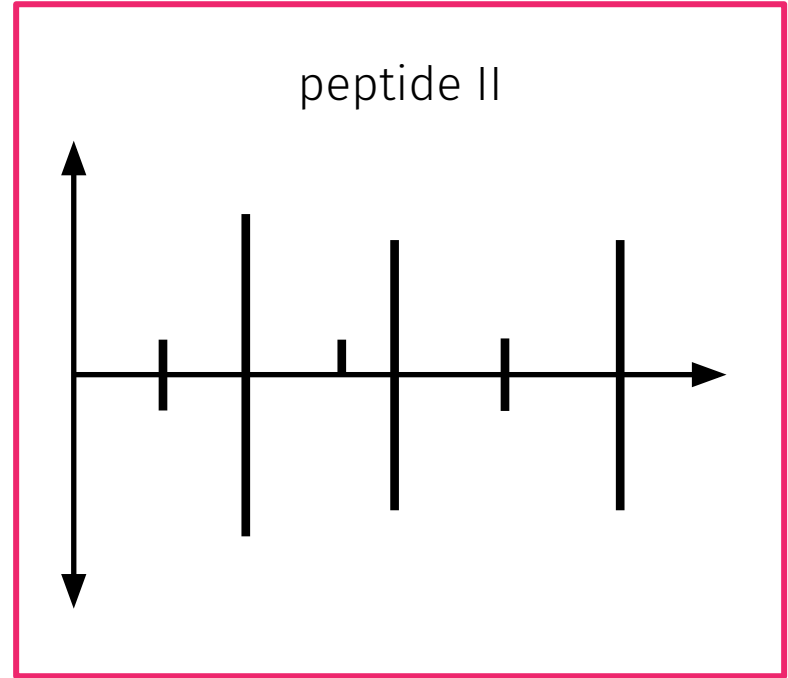
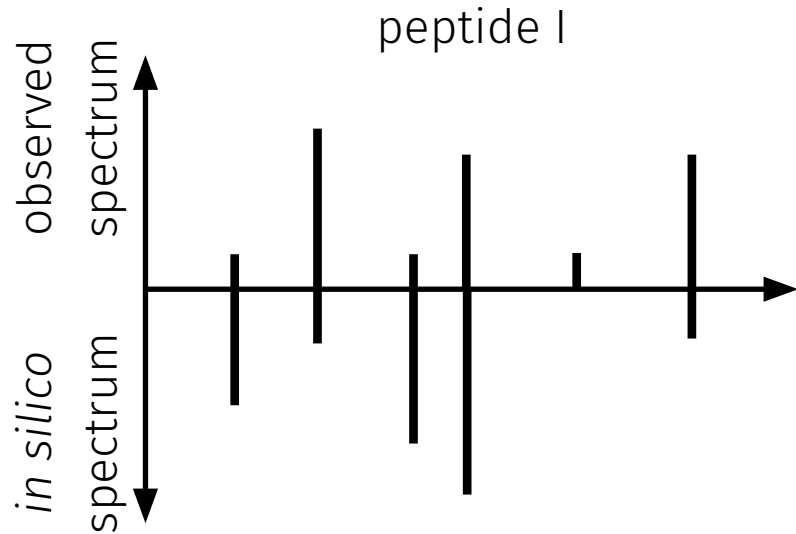
SPECTRUM ANNOTATION IS CHALLENGING



INTENSITY INFORMATION AVOIDS FALSE POSITIVES

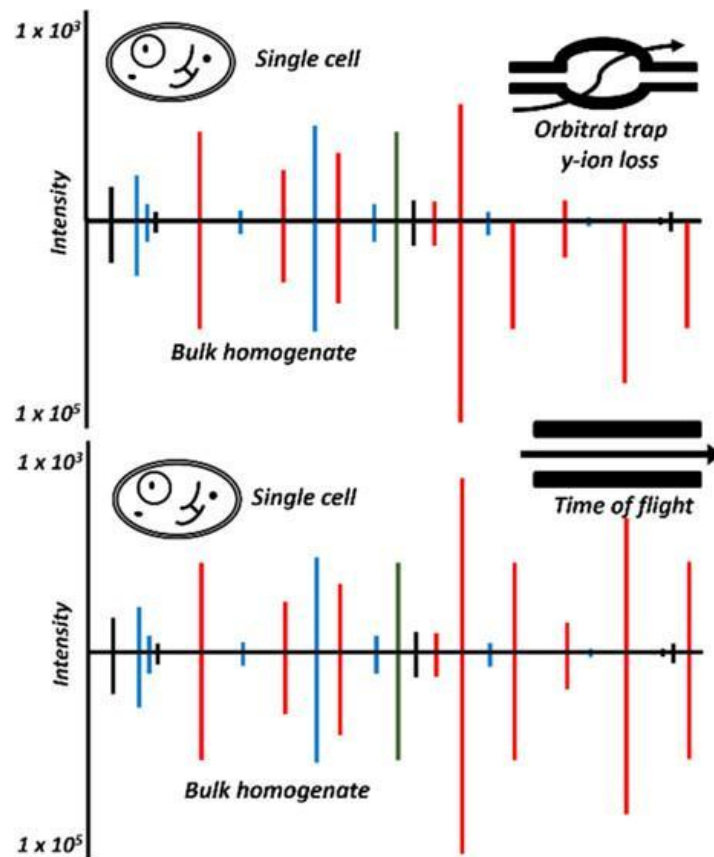


INTENSITY INFORMATION AVOIDS FALSE POSITIVES



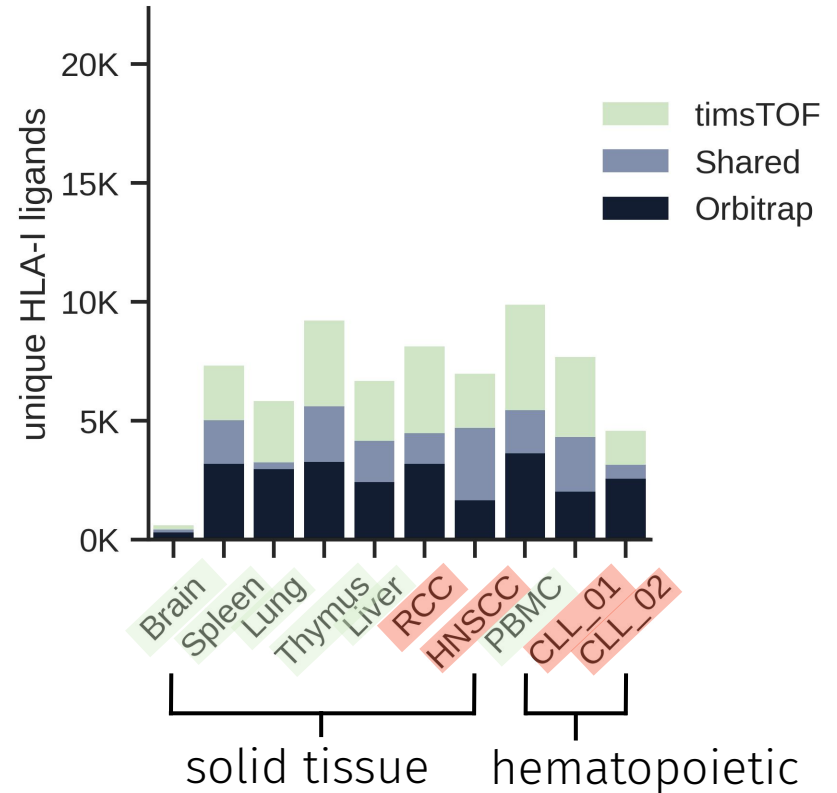
VALUE OF TIMSTOF FOR IMMUNOPEPTIDOMICS

- timsTOF stays stable at low abundances
- A few immunopeptides can elicit an immune response

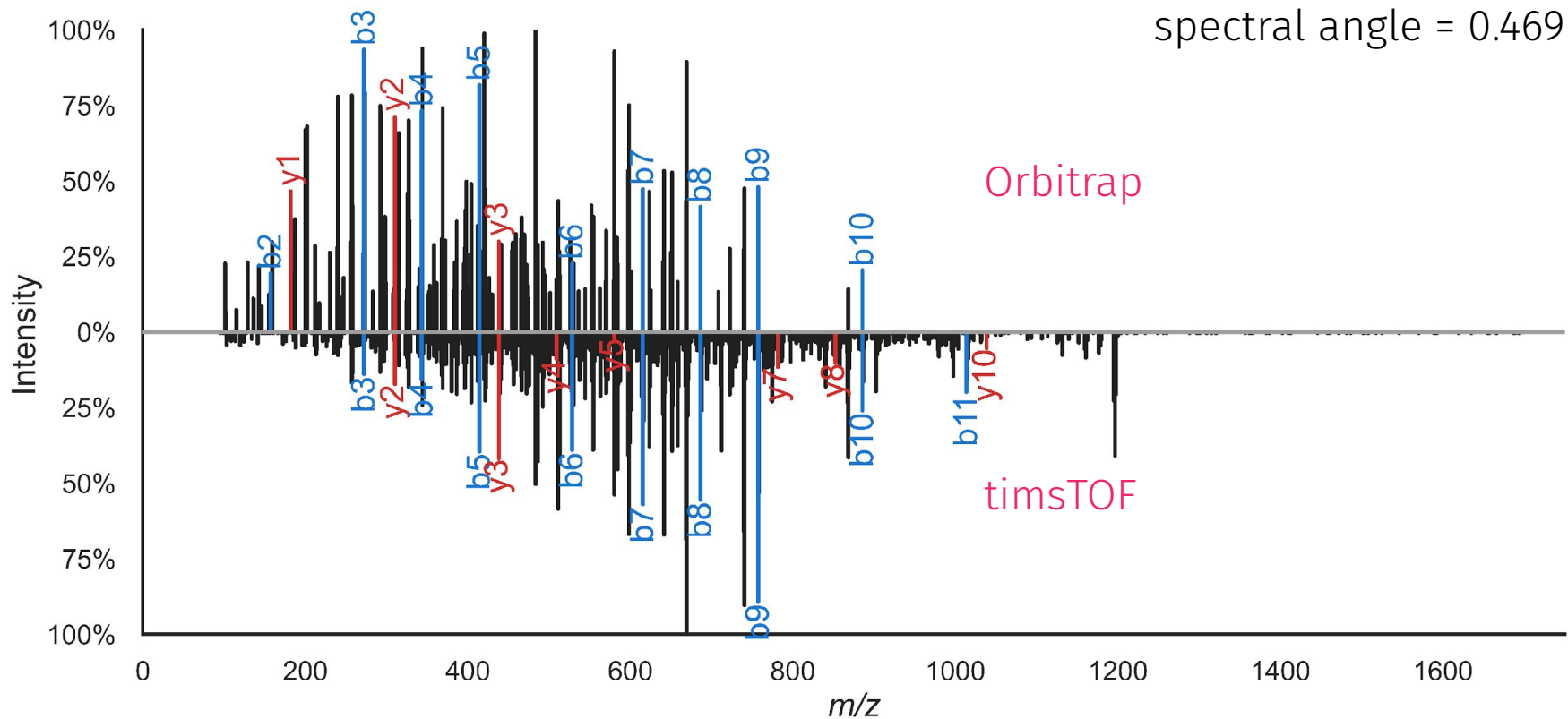


1.5-FOLD PEPTIDE INCREASE ON TIMSTOF

- Benign and malignant samples
- Measured on both Orbitrap and timsTOF



NEED FOR A TIMSTOF INTENSITY PREDICTION MODEL

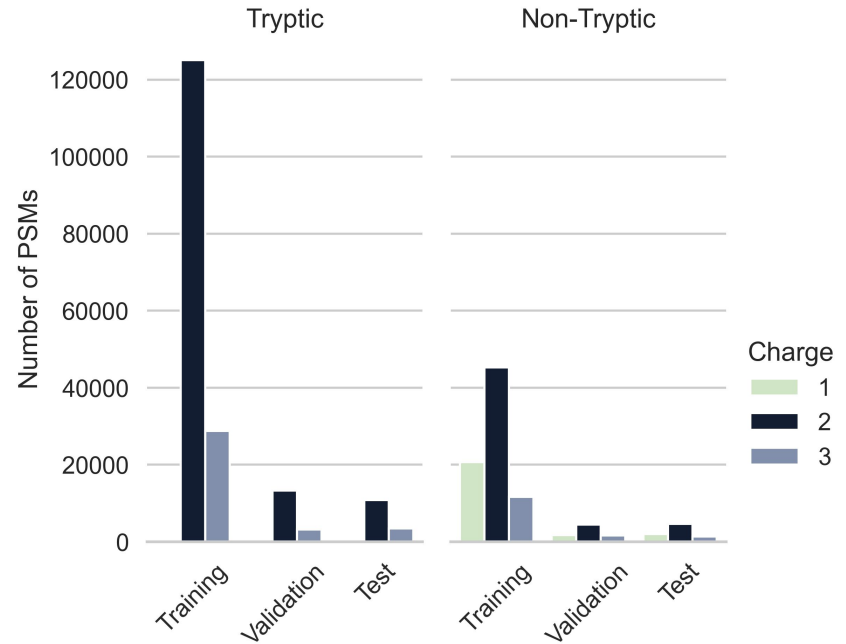


ORIGIN OF THE TRAINING DATA



Kurt Boonen

- Measured >300,000 non-tryptic synthesized peptides
- >120,000 previously acquired tryptic synthesized peptides

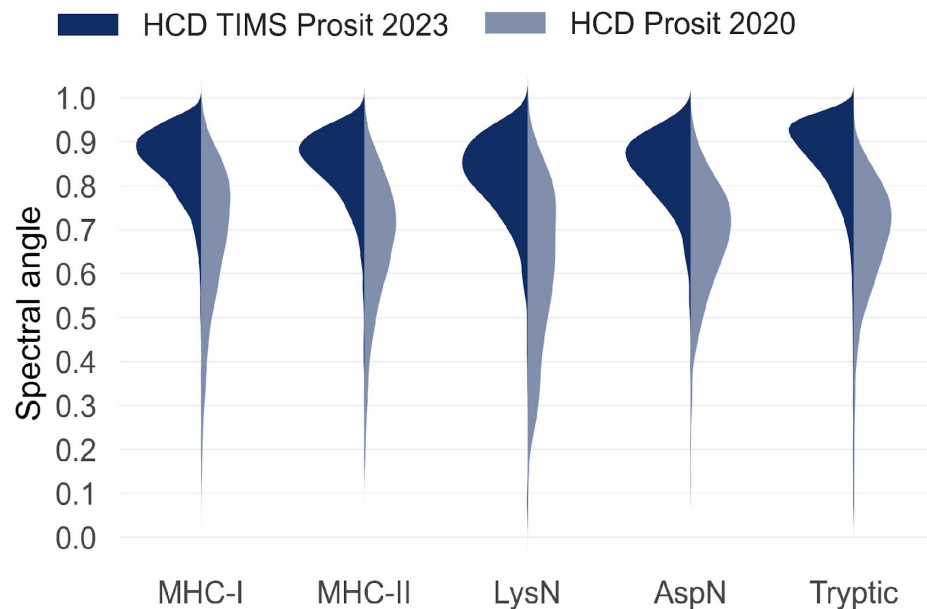


FINE-TUNING THE PROSIT HCD MODEL

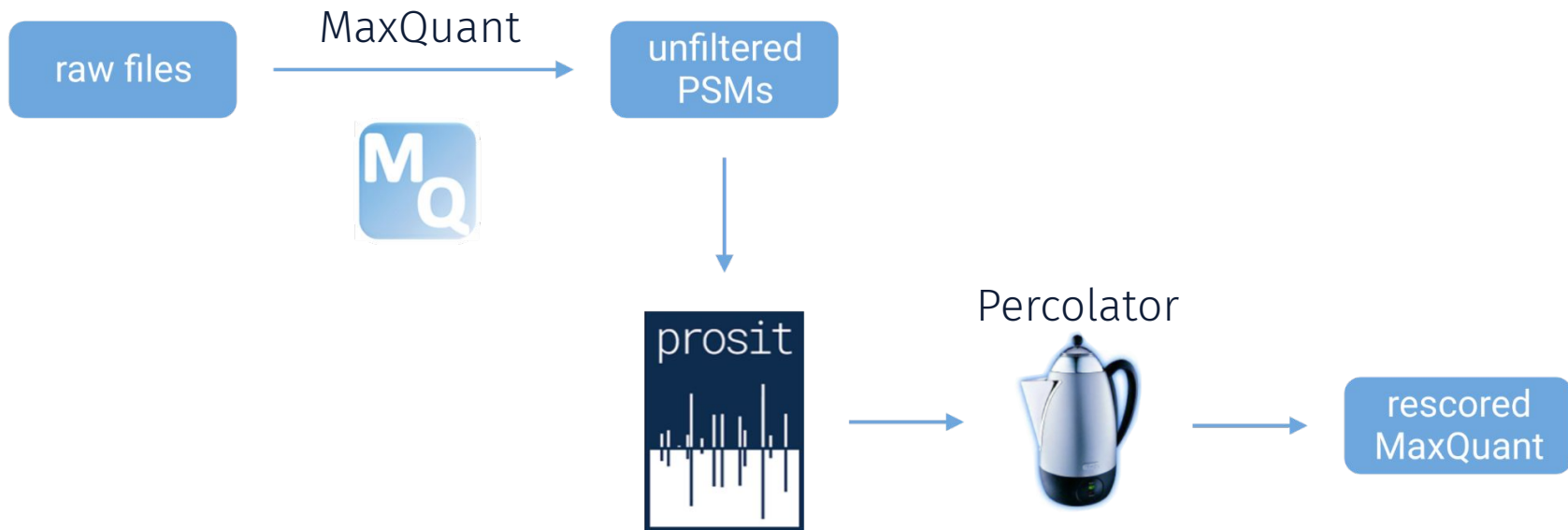


Charlotte Adams

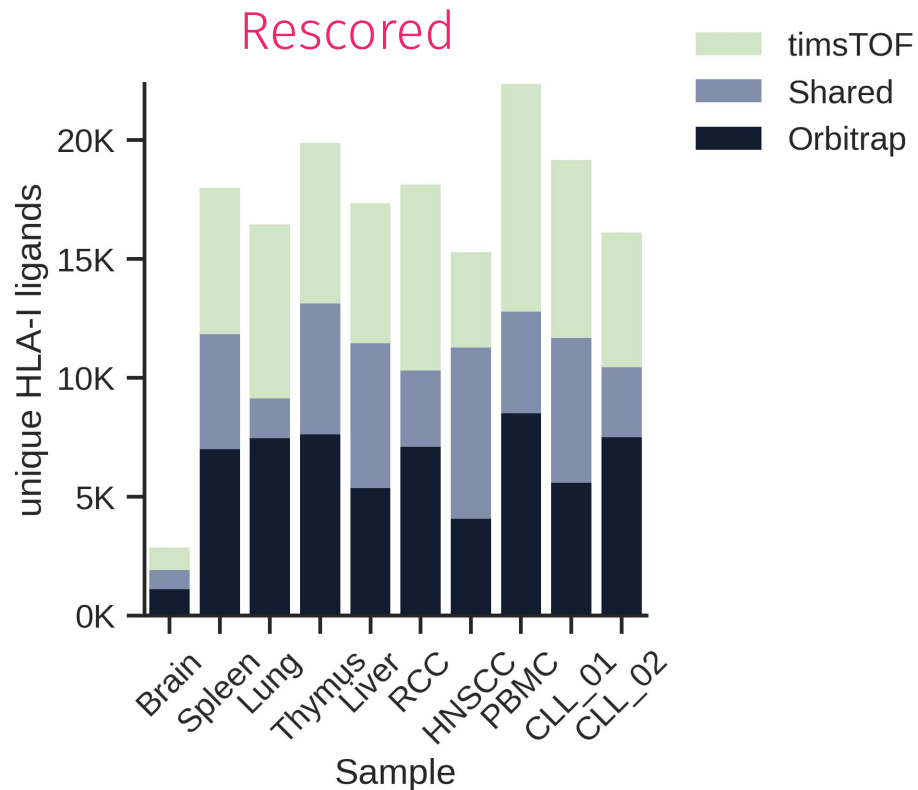
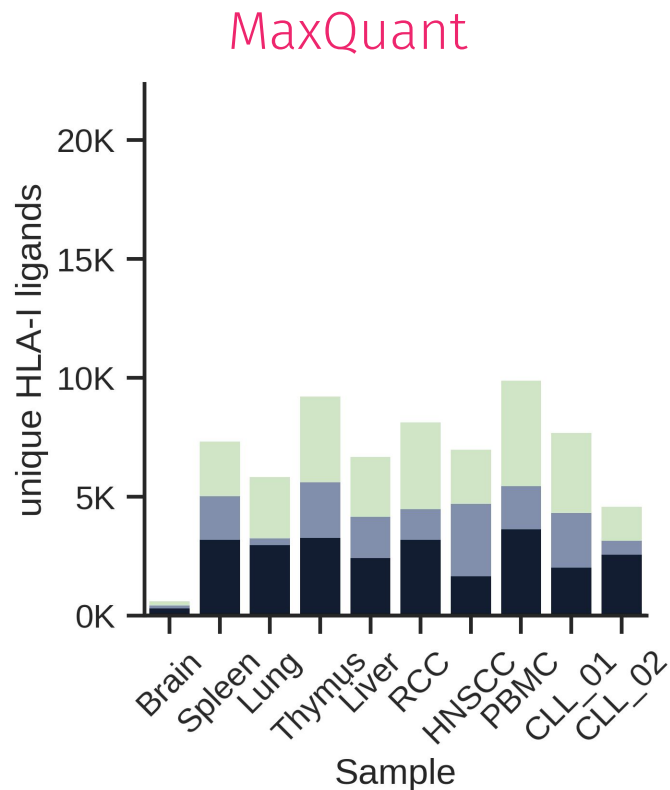
- Prosit HCD 2020 was trained on ~30 million spectra (9 million non-tryptic spectra)
- ~280,000 timsTOF spectra
- Improved spectral angle between predicted and experimental spectra



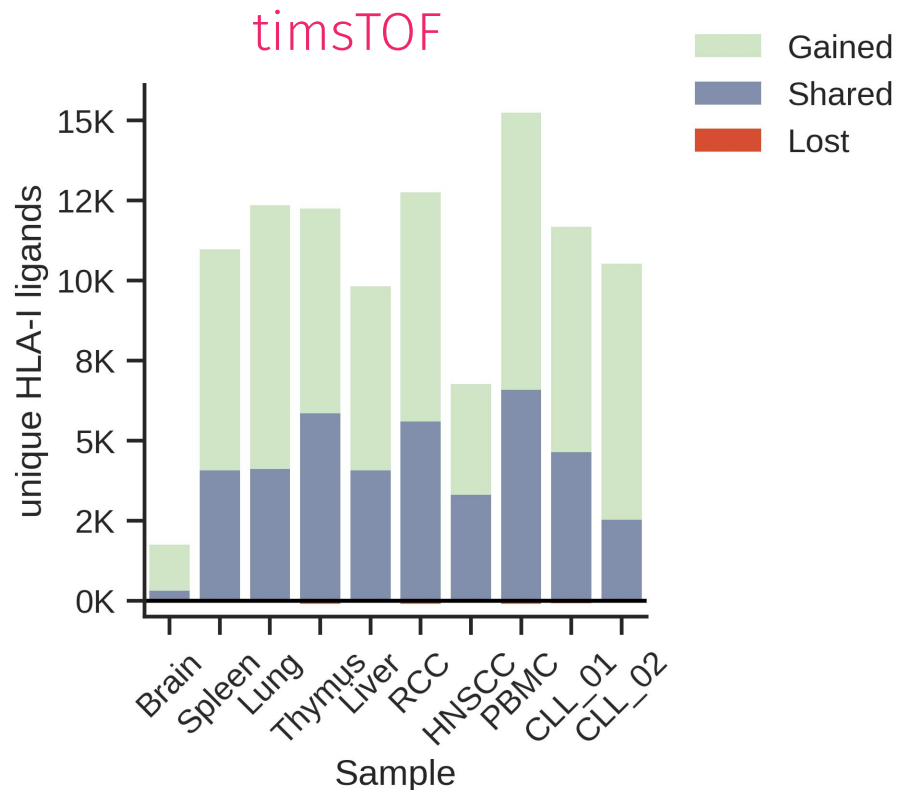
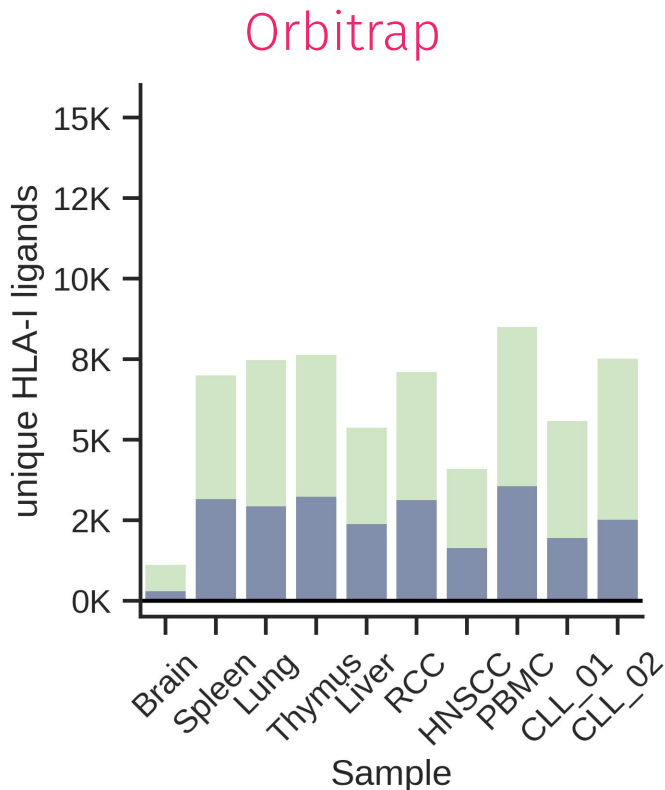
OUR RESCORING PIPELINE



2.4-FOLD PEPTIDE INCREASE AFTER RESCORING



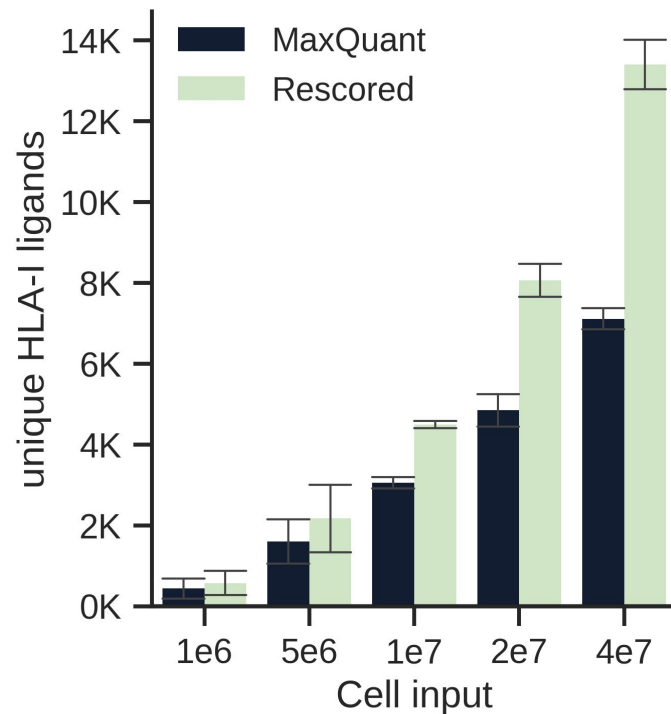
INCREASED PERFORMANCE ON TIMSTOF VS ORBITRAP



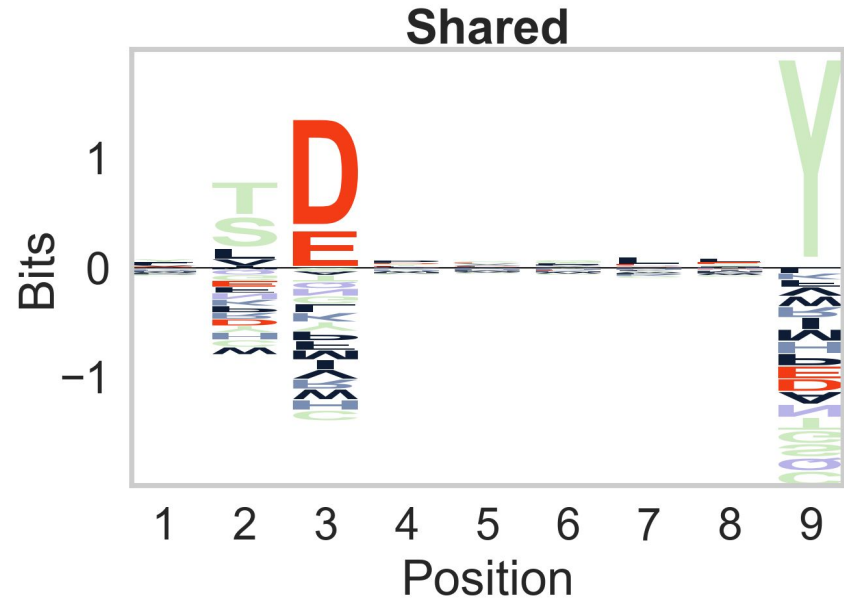
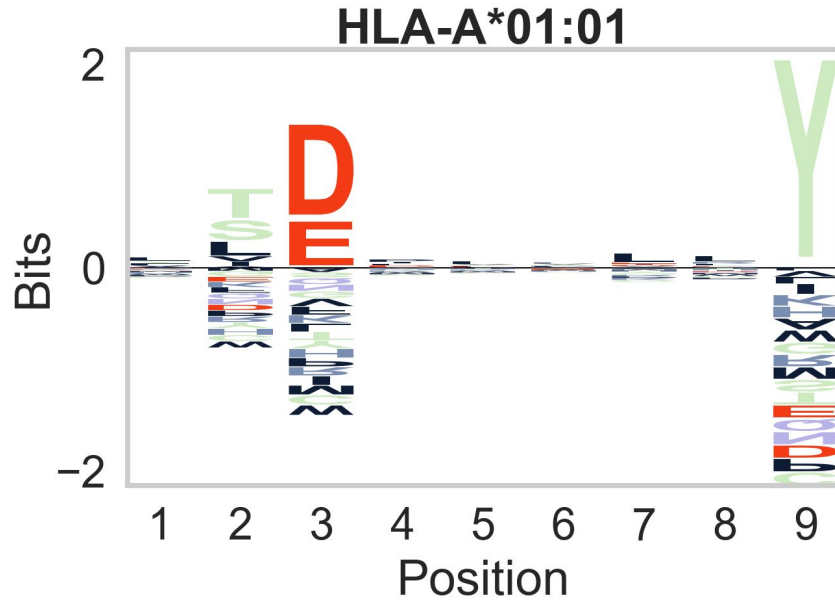
IMMUNOPEPTIDOMICS ON MELANOMA CELLS

- A375, a melanoma cell line
- Measured in triplicate
- Missense SNPs were added to the FASTA file
- HLA types

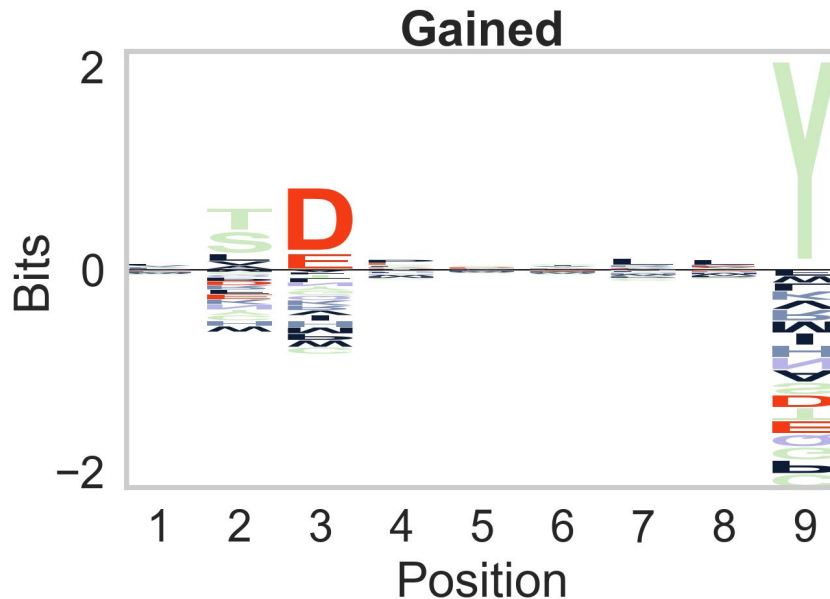
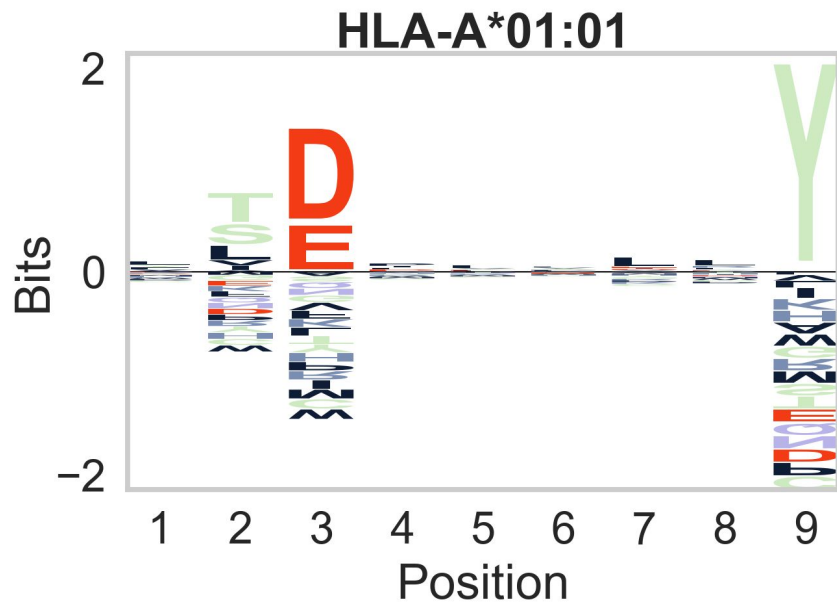
HLA-A	HLA-B	HLA-C
01:01 + 02:02	57:01 + 44:03	16:02 + 06:02



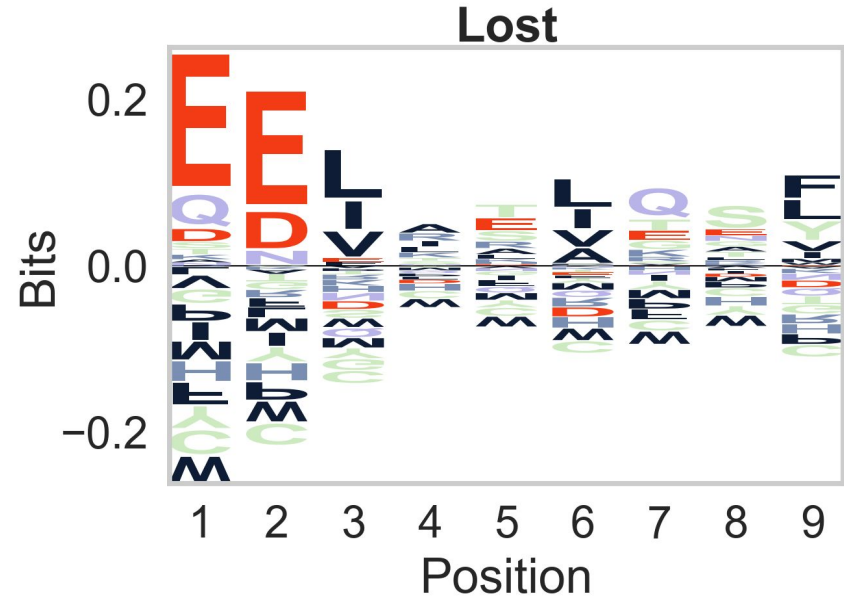
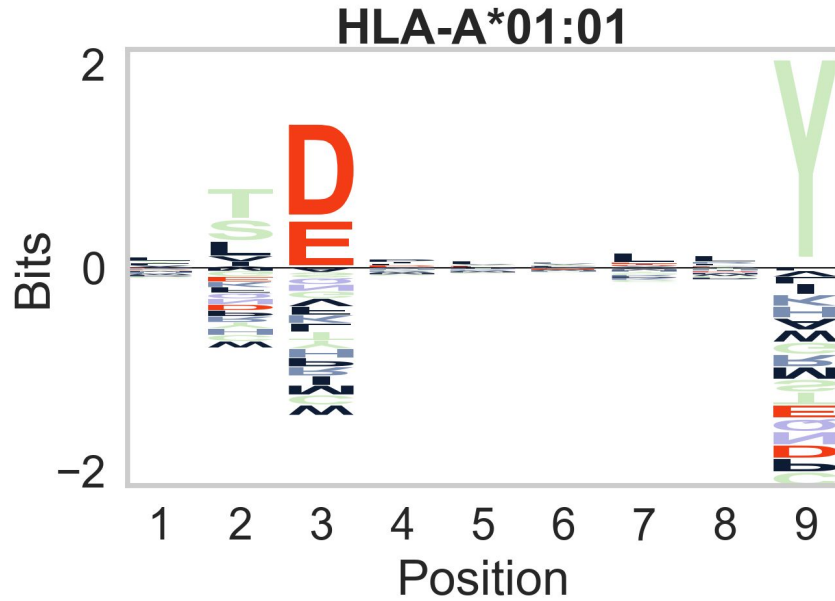
MOTIFS OF IDENTIFIED PEPTIDES MATCH HLA TYPES



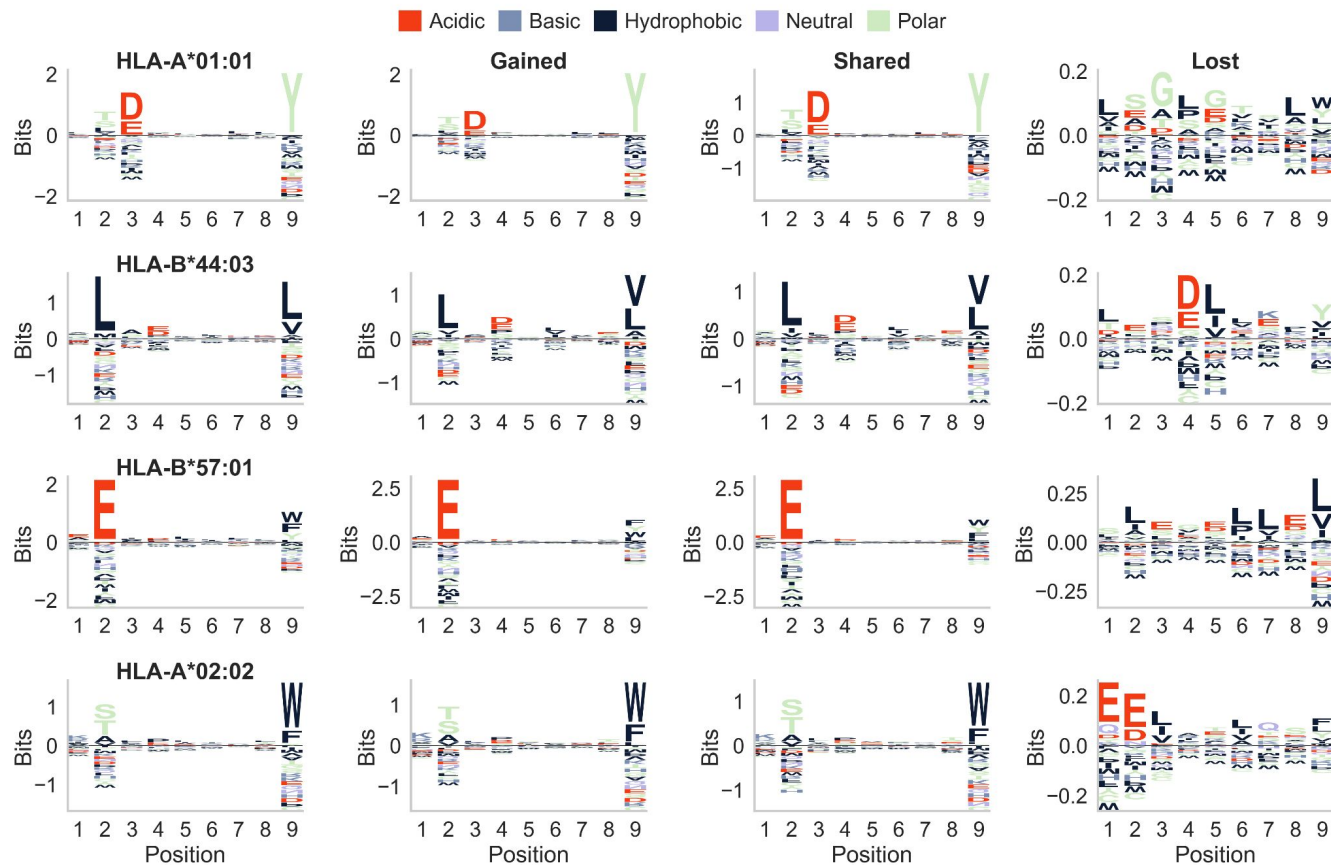
MOTIFS OF IDENTIFIED PEPTIDES MATCH HLA TYPES



MOTIFS OF IDENTIFIED PEPTIDES MATCH HLA TYPES

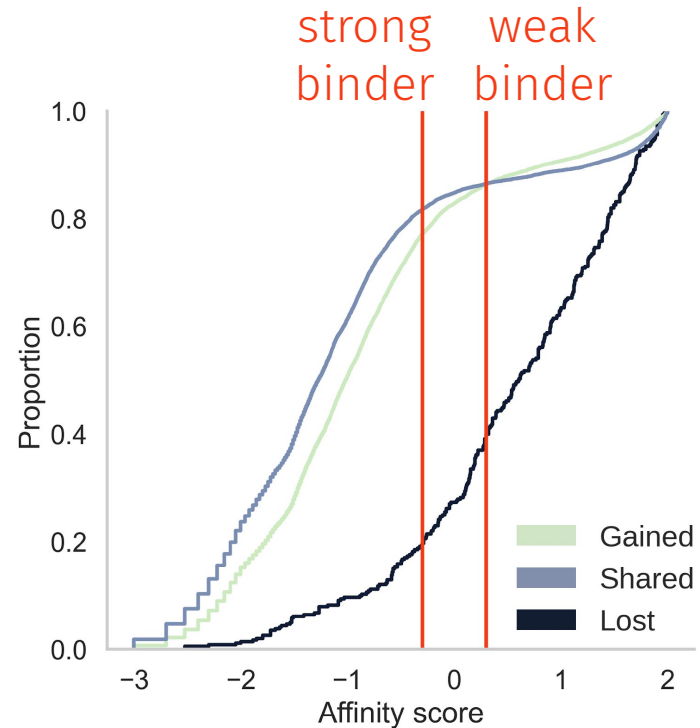


MOTIFS OF IDENTIFIED PEPTIDES MATCH HLA TYPES



IDENTIFIED PEPTIDES HAVE STRONG HLA BINDING

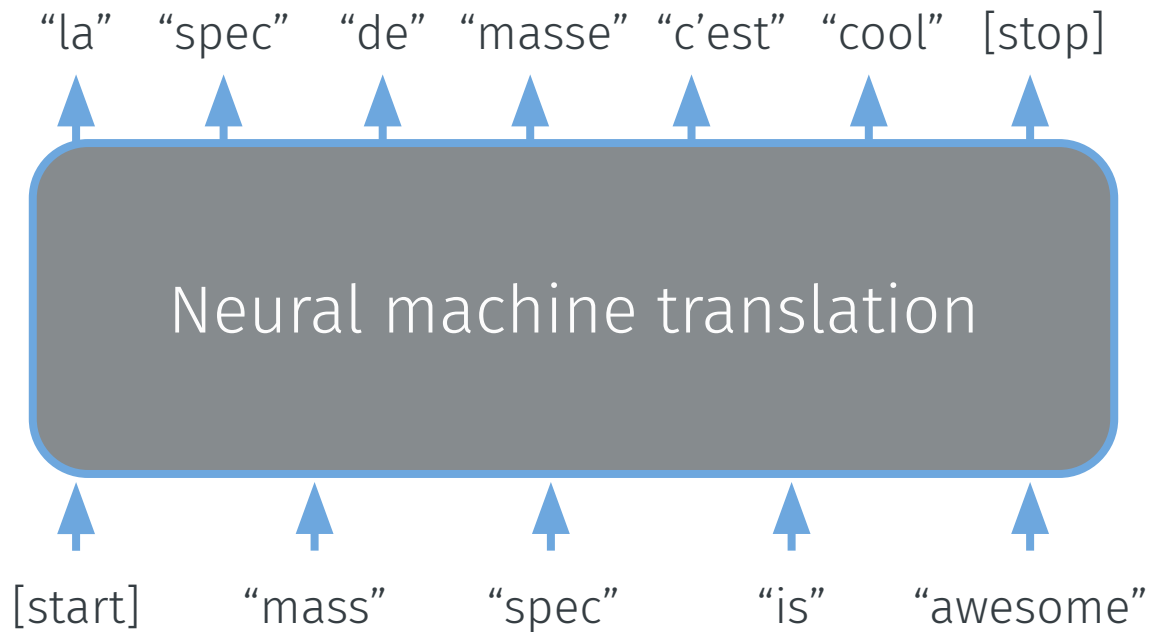
- NetMHCpan to predict the HLA binding affinity
- Best (=lowest) score selected for each peptide against HLA types
- 86% of peptides after rescoring are at least weak binders





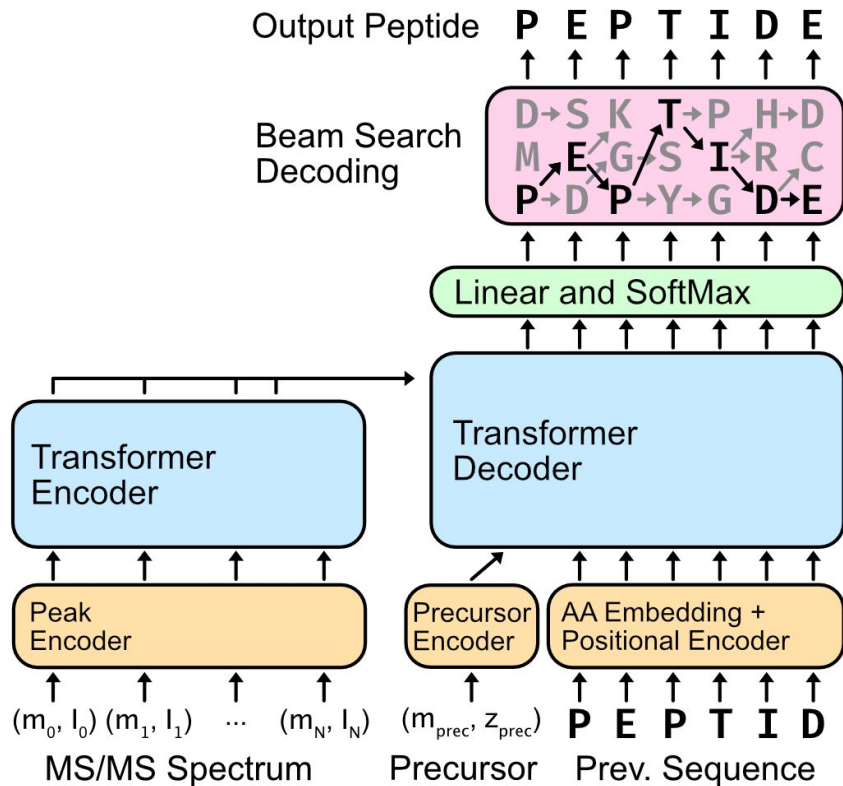
**SOLUTION 2:
DE NOVO PEPTIDE SEQUENCING**

PEPTIDE SEQUENCING AS A TRANSLATION PROBLEM

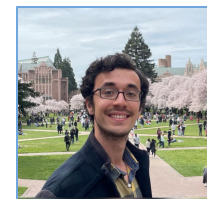


CASANOVO IS A *DE NOVO* SEQUENCING TRANSFORMER

- **Encoder:** Learns contextualized representations for peaks
- **Decoder:** Predicts the *de novo* sequence one amino acid at a time
- **Beam search** decoding finds the highest-scoring peptide

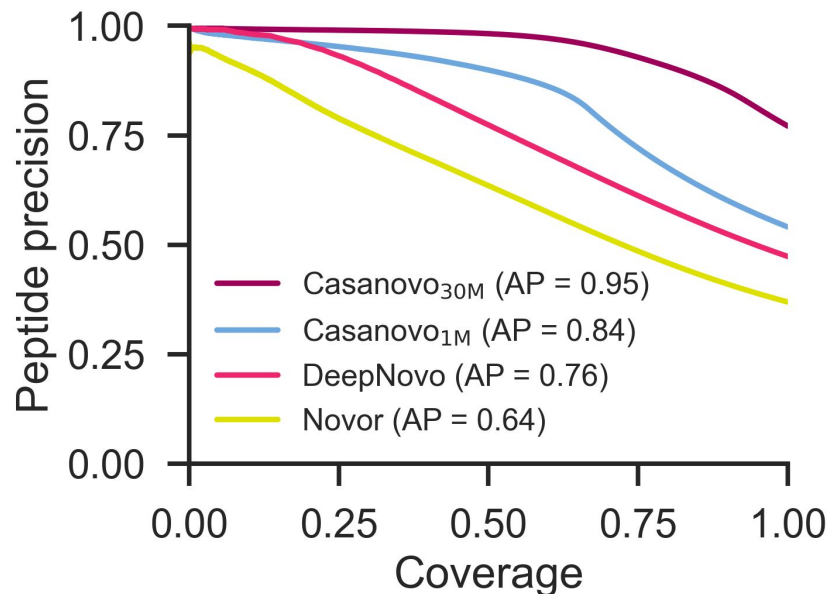


CASANOVO OUTPERFORMS OTHER METHODS



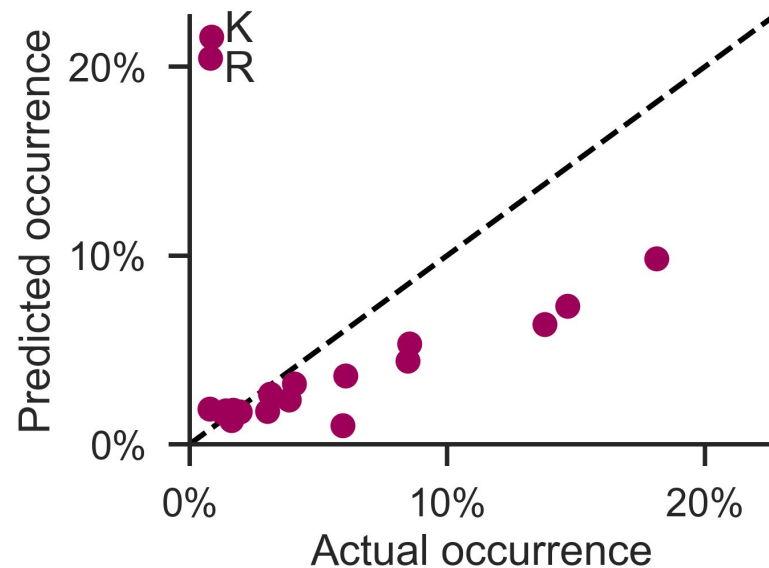
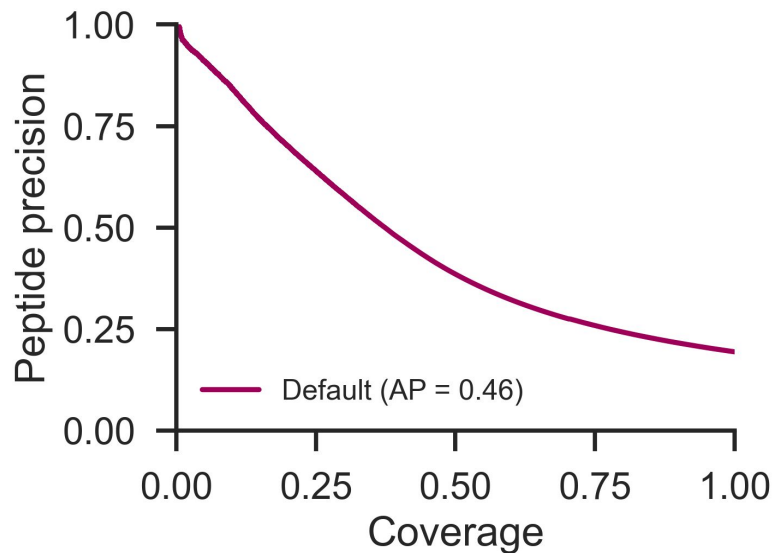
Melih Yilmaz

- With same training data, Casanovo outperforms traditional and deep learning-based *de novo* sequencing tools
- Large training data significantly improves performance



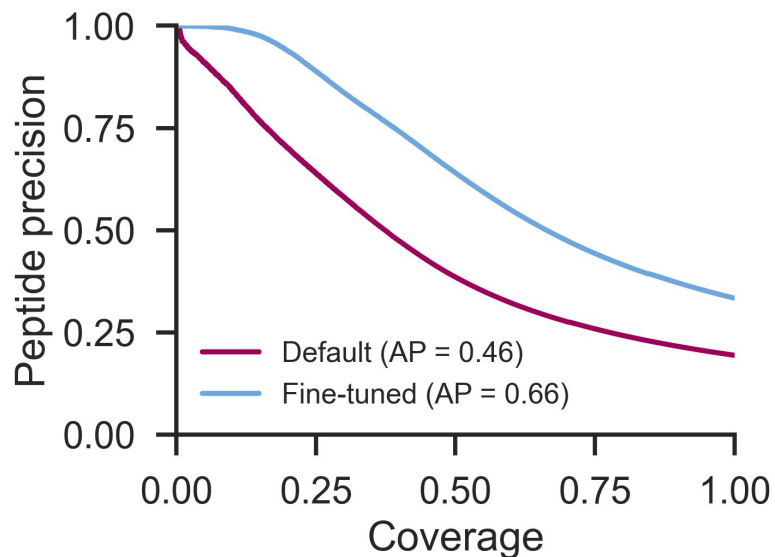
NON-ENZYMATIC FINE-TUNING AVOIDS TRYPTIC BIAS

- Casanovo is trained on bottom-up proteomics data, exhibiting a tryptic bias

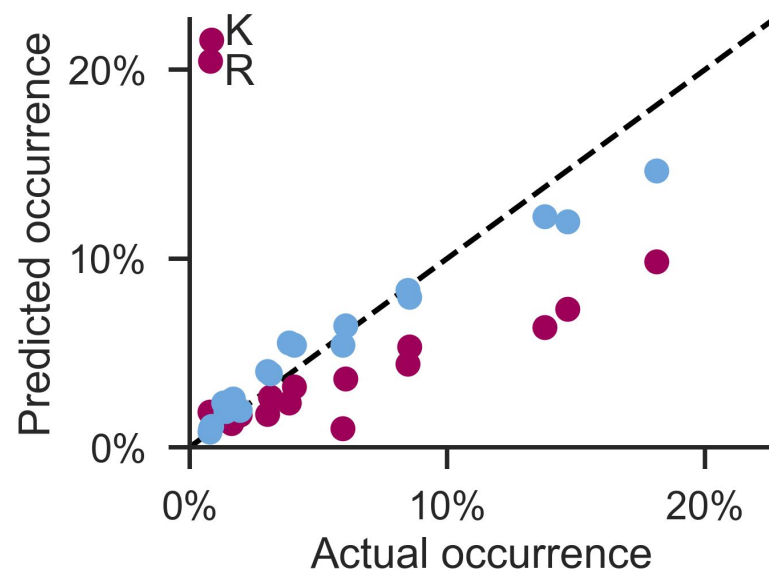


NON-ENZYMATIC FINE-TUNING AVOIDS TRYPTIC BIAS

- Casanovo is trained on bottom-up proteomics data, exhibiting a tryptic bias



- Fine-tuning on non-enzymatic data mitigates the tryptic bias and improves performance

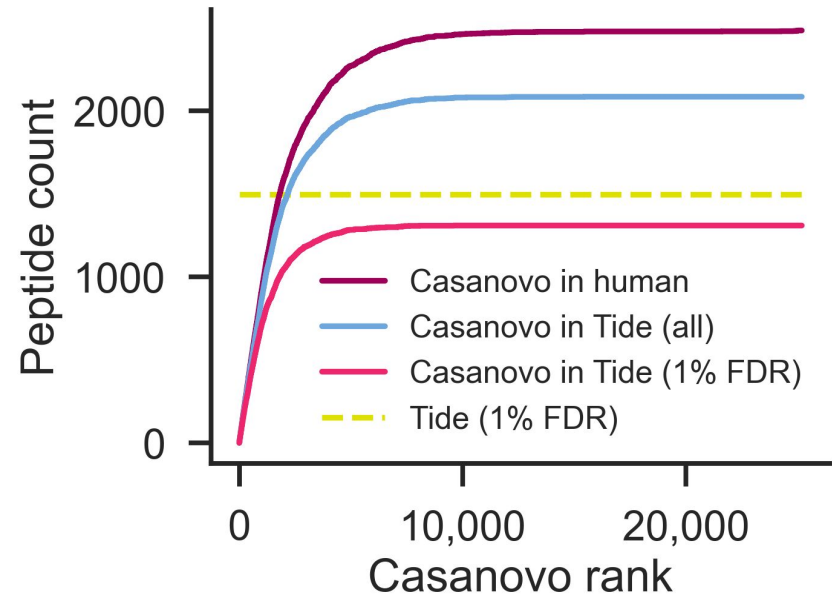


CASANOVO OUTPERFORMS DATABASE SEARCHING

Check if Casanovo predictions are:

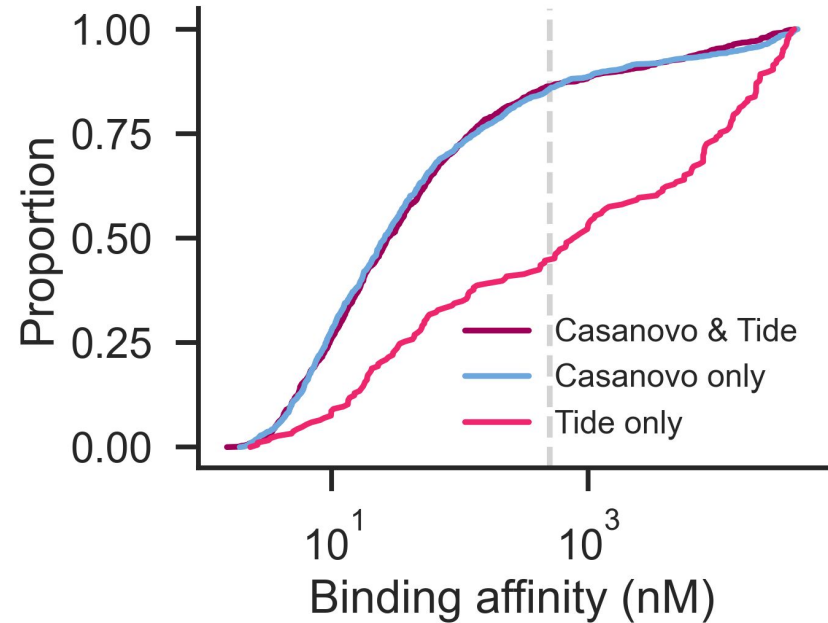
- In the human proteome
- Database search results (Tide) on human proteome

→ Casanovo identifies 65% more unique peptides matching to the human proteome than Tide

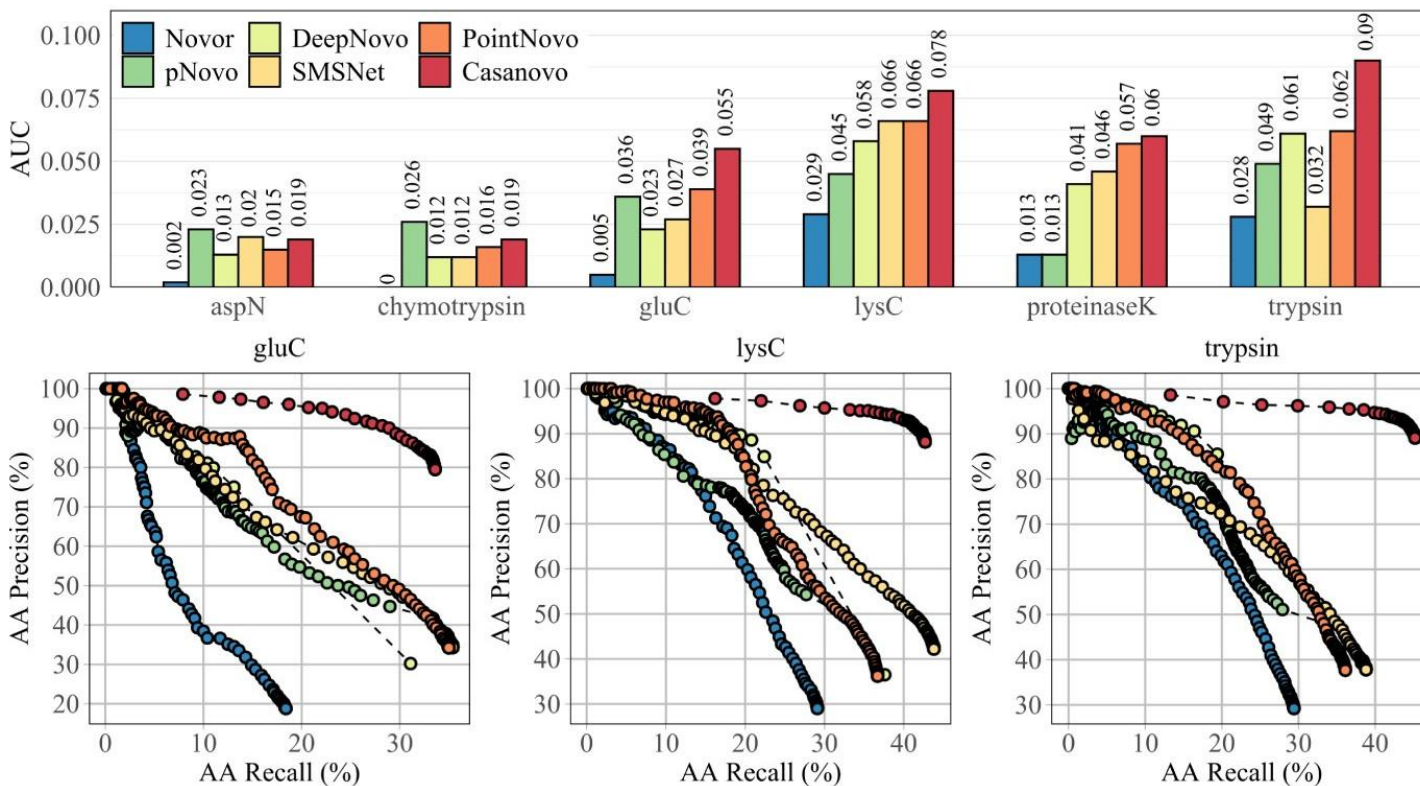


IDENTIFIED PEPTIDES HAVE STRONG HLA BINDING

- NetMHCpan to predict the HLA binding affinity
- Casanovo predicted peptides in human are:
 - As plausible as Casanovo & Tide shared peptides
 - More plausible than peptides uniquely identified by Tide

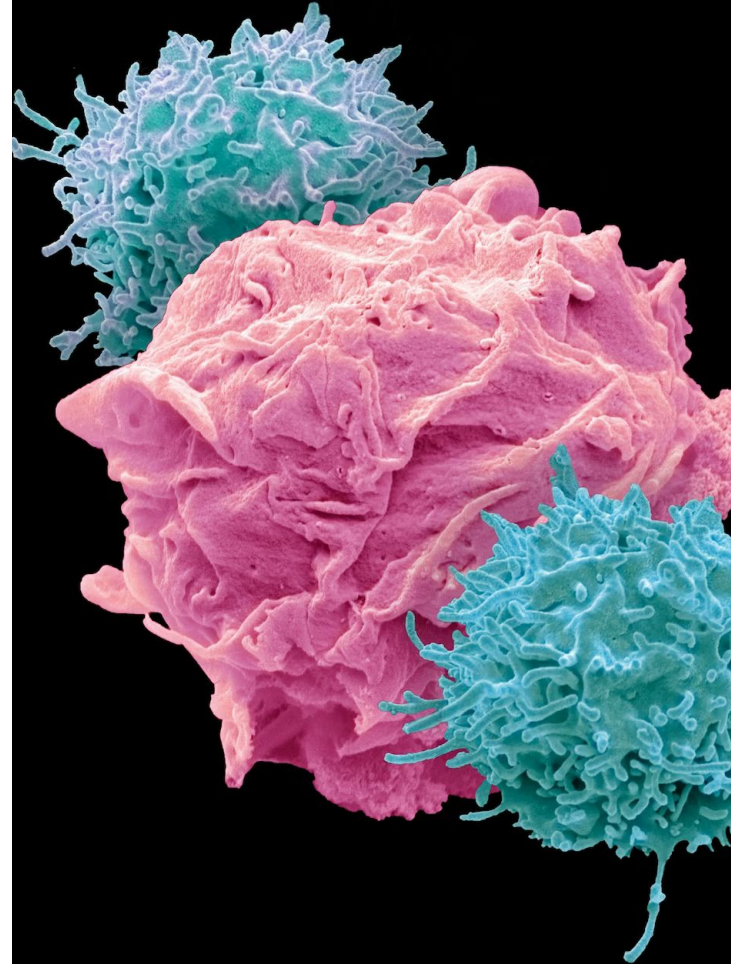


EXTERNAL EVALUATION OF ANTIBODY ASSEMBLY



CONCLUSION

- Identifying immunopeptides is challenging
- Fragment ion intensity prediction for rescoring on timsTOF data
- Casanovo is a powerful deep learning-based *de novo* peptide sequencing solution



ACKNOWLEDGEMENTS

University of Antwerp

Charlotte Adams

Kurt Boonen

Kris Laukens

Technical University of Munich

Mathias Wilhelm

Wassim Gabriel

University of Washington

Bill Noble

Melih Yilmaz

Will Fondrie

Rowan Nelson

Varun Ananth

Sewoong Oh



Universiteit
Antwerpen

