

MACHINE LEARNING APPLICATIONS FOR IMMUNOPEPTIDOMICS ANALYSIS

Wout Bittremieux — HUPO 2023 Bioinformatics Hub



@wout@sigmoid.social

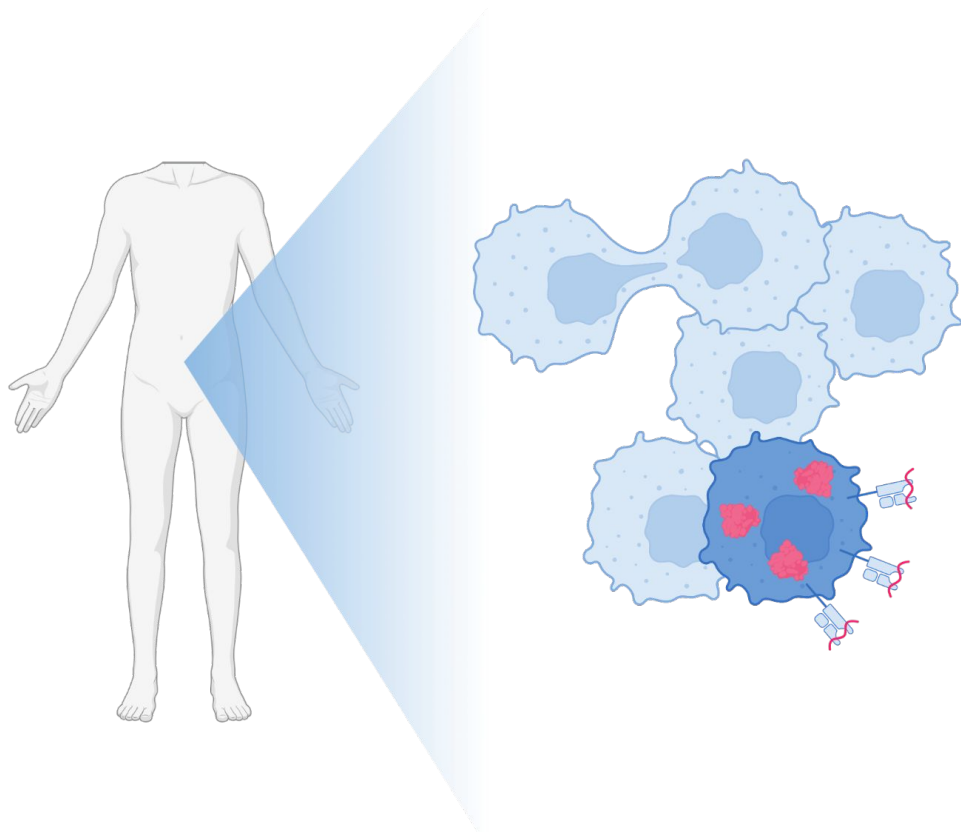


bittremieux.be

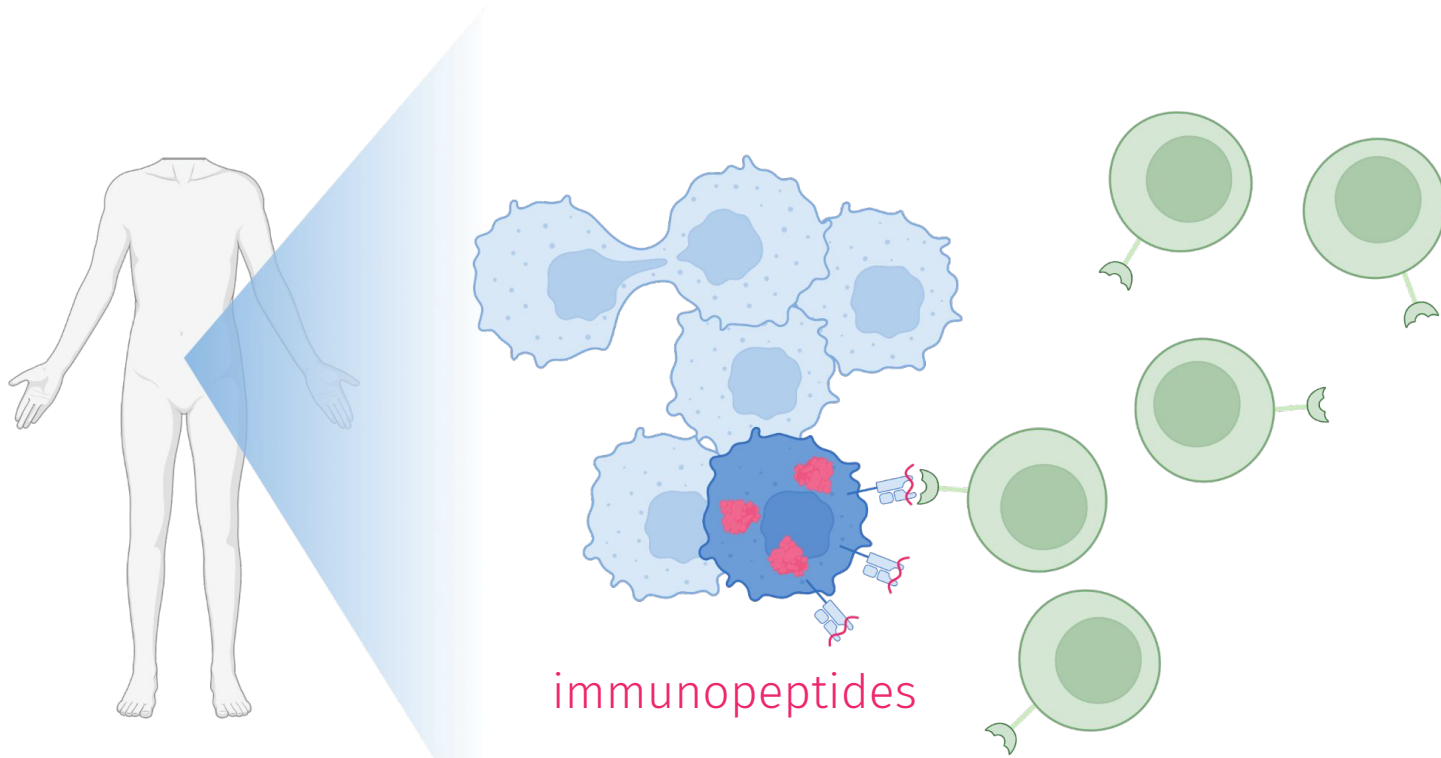


wout.bittremieux@uantwerpen.be

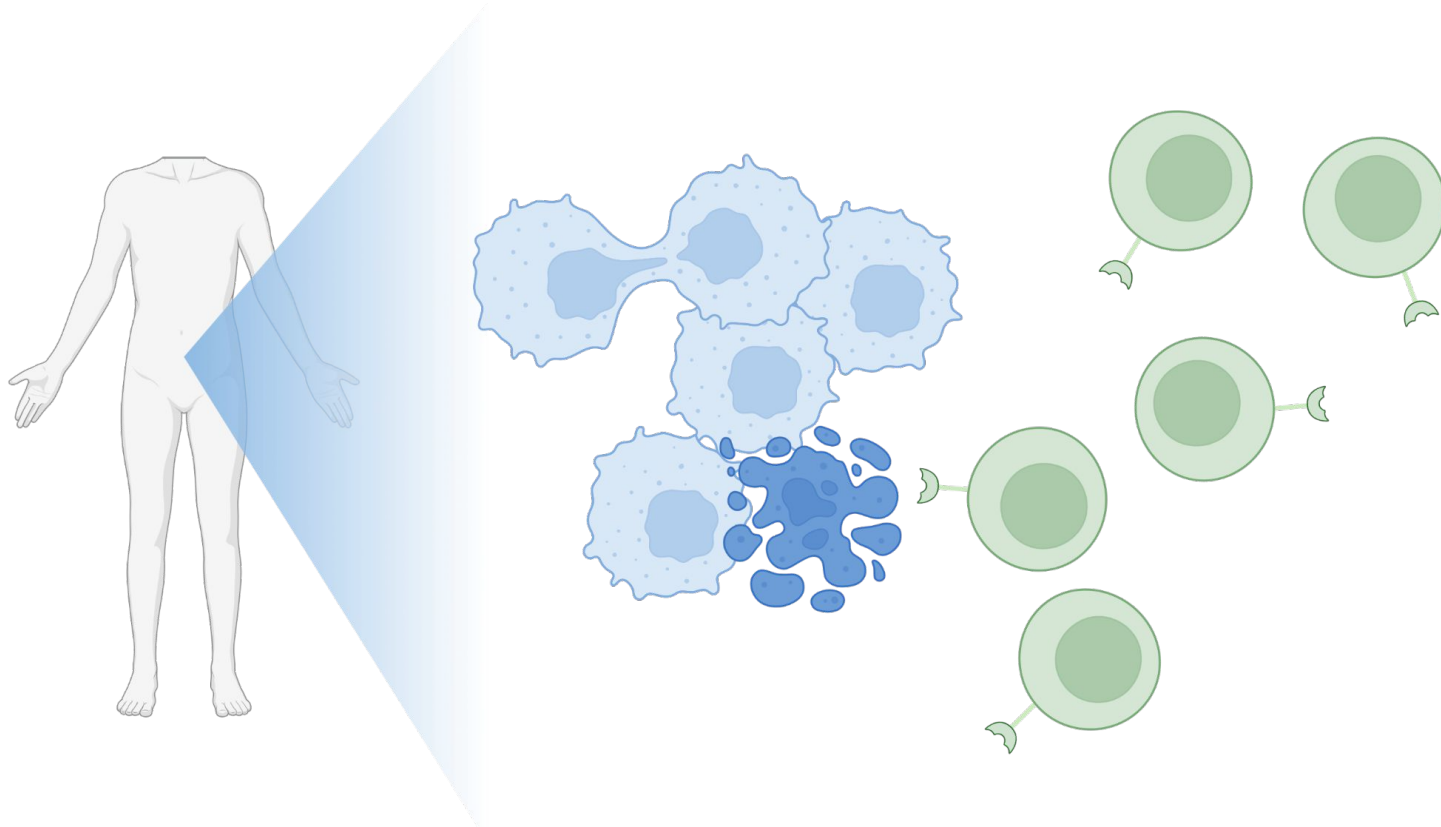
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



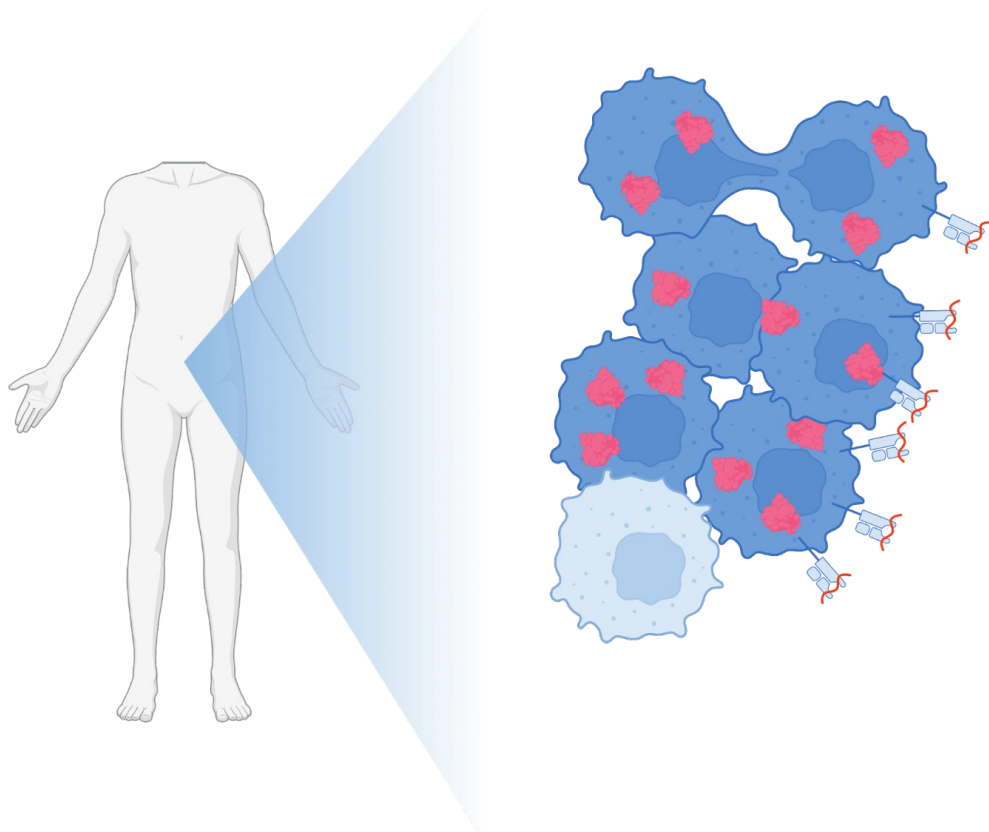
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



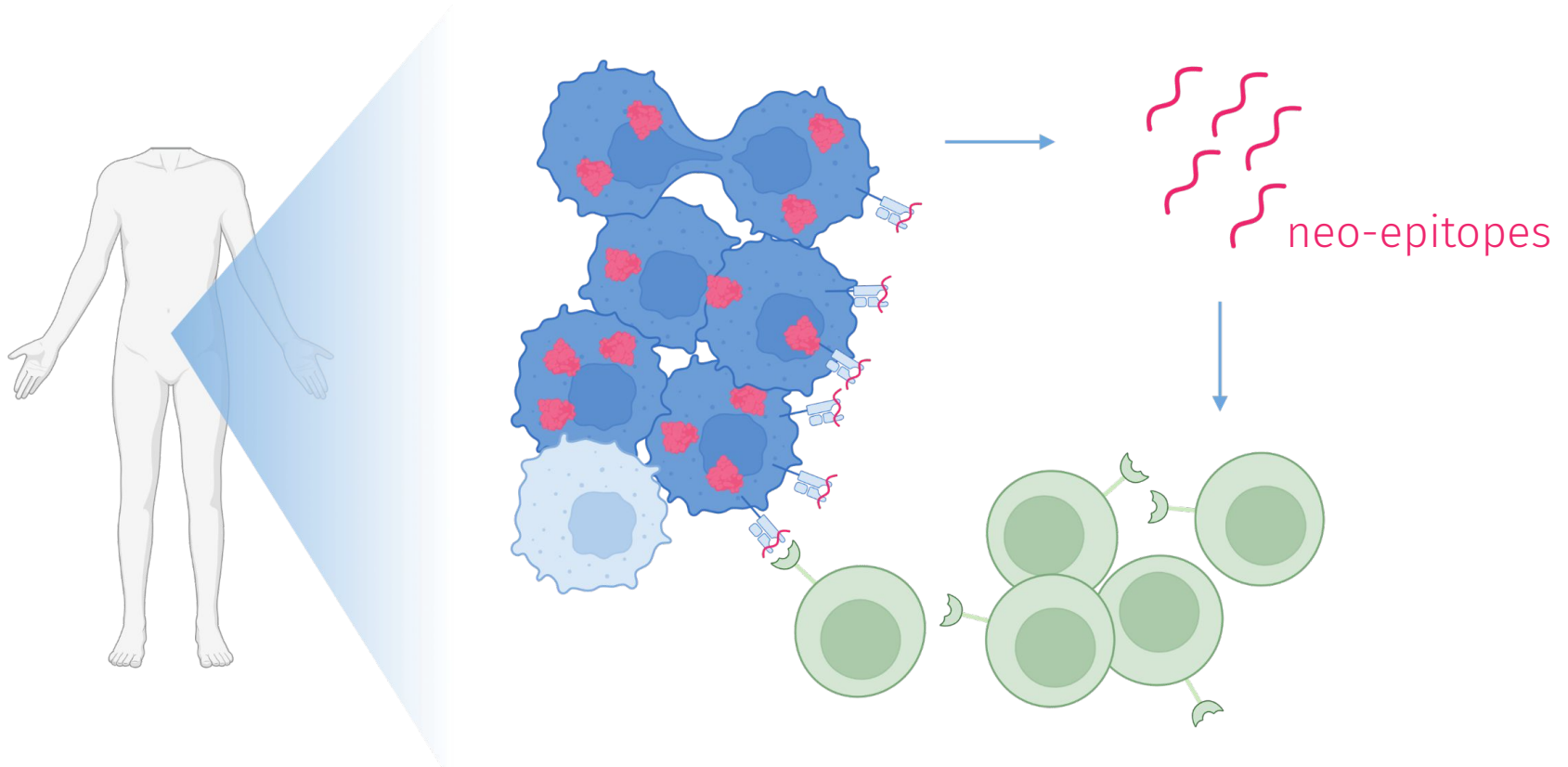
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



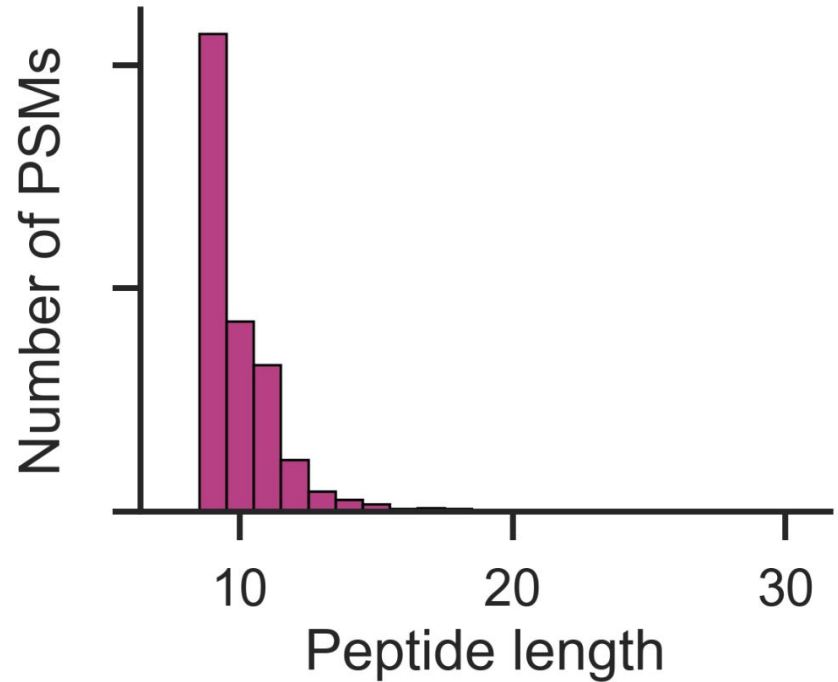
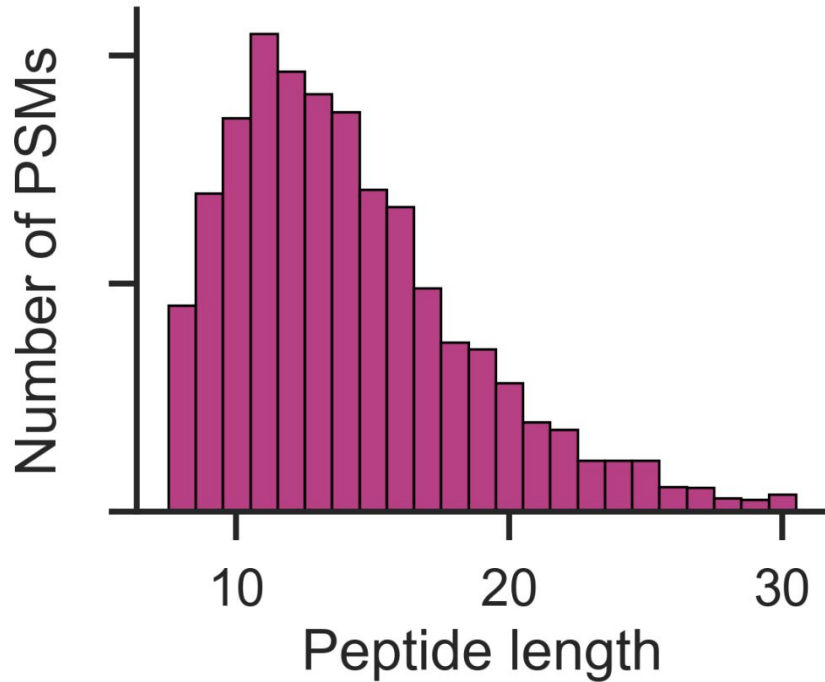
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



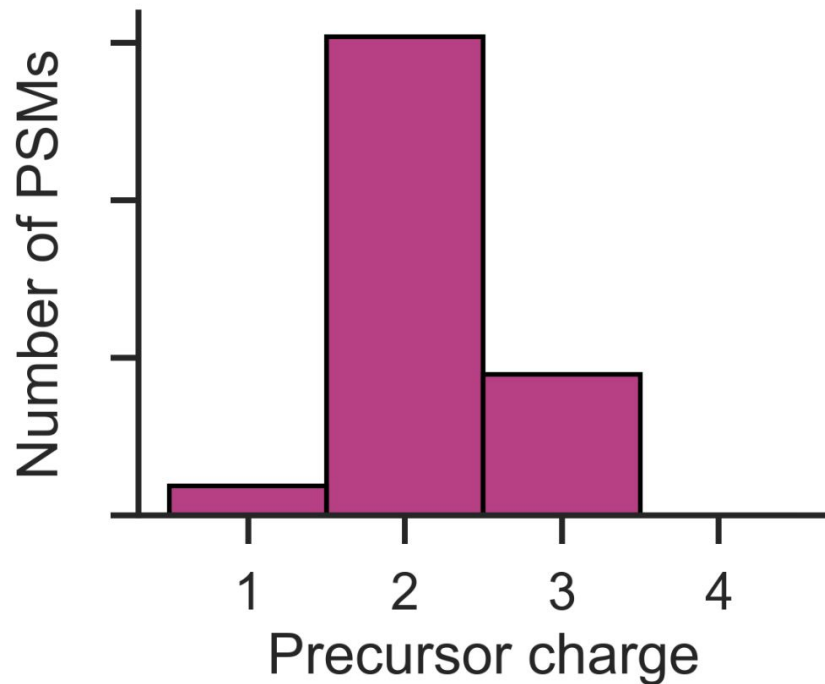
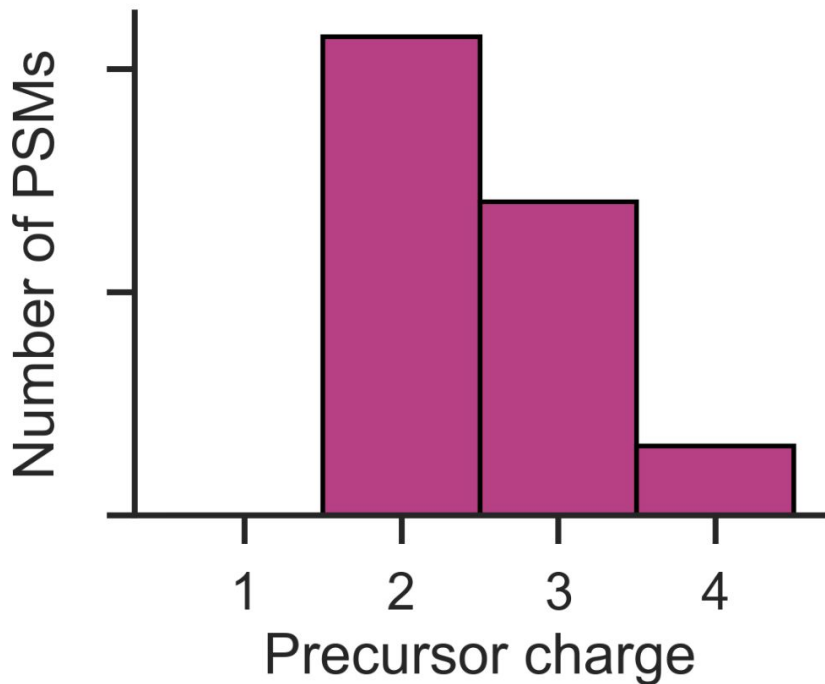
WHY INVESTIGATE THE IMMUNOPEPTIDOME?



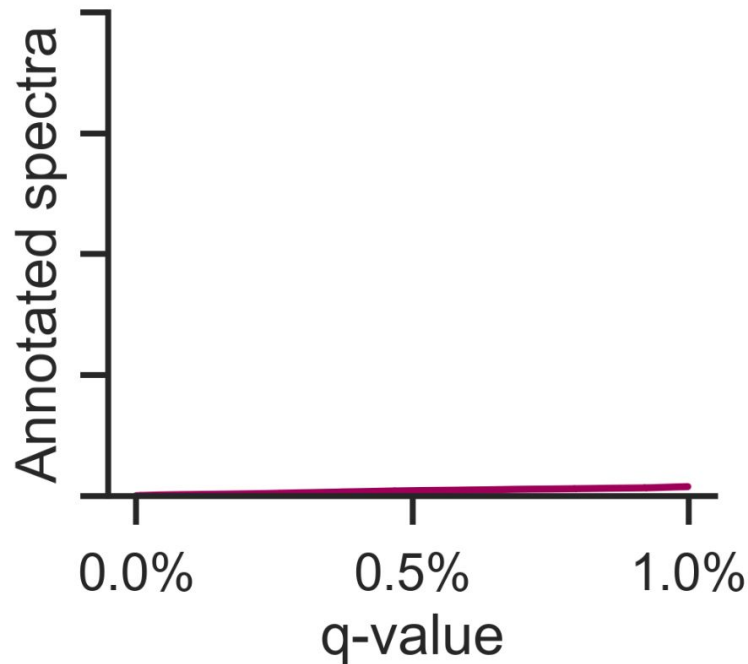
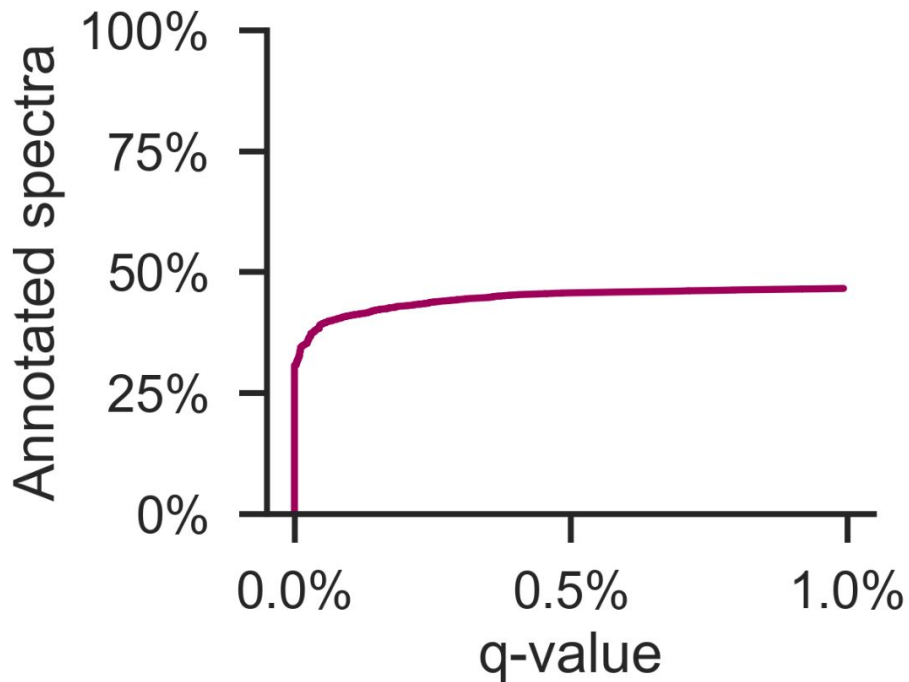
QUIZ: BOTTOM-UP OR IMMUNOPEPTIDOMICS?



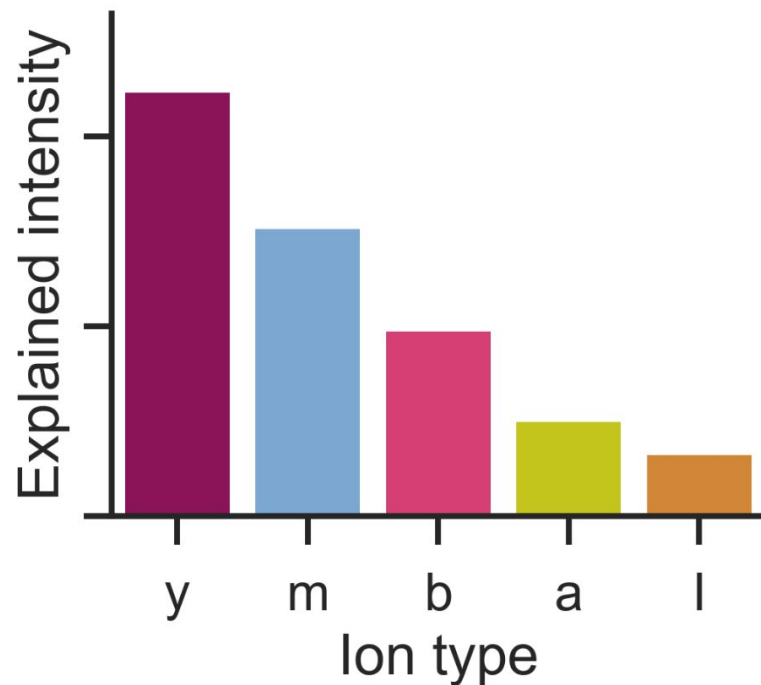
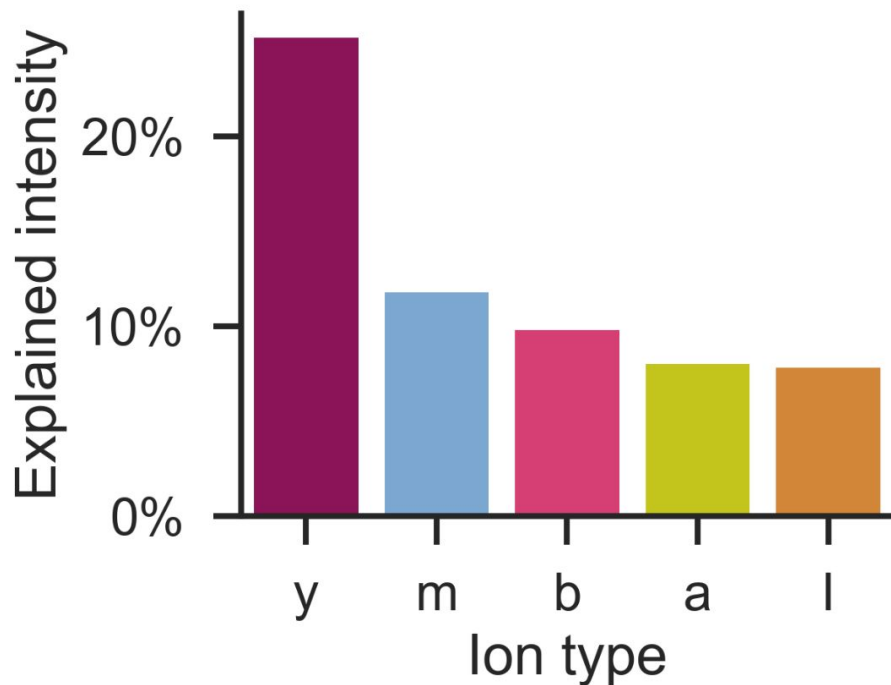
QUIZ: BOTTOM-UP OR IMMUNOPEPTIDOMICS?



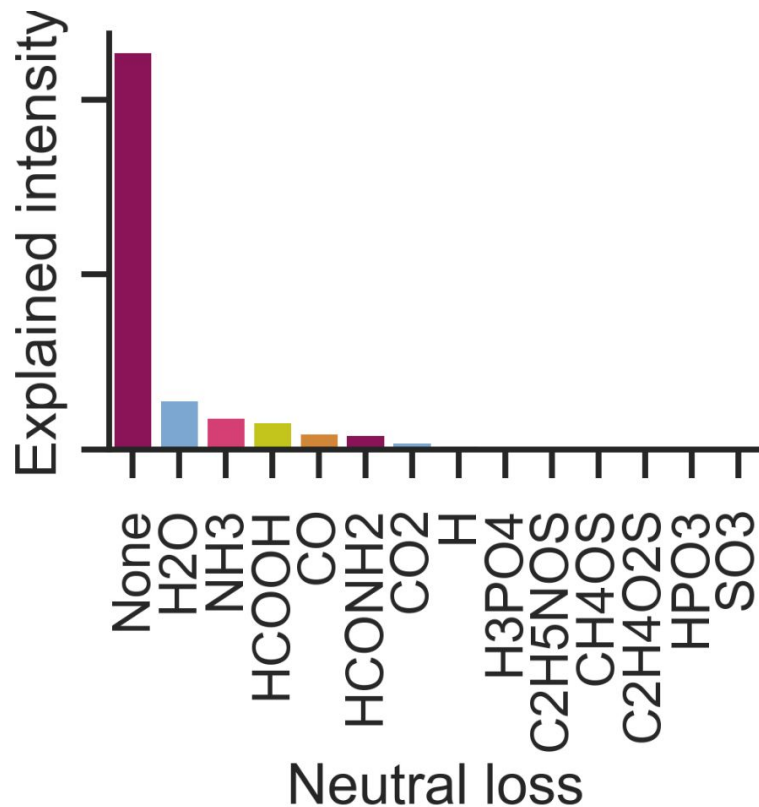
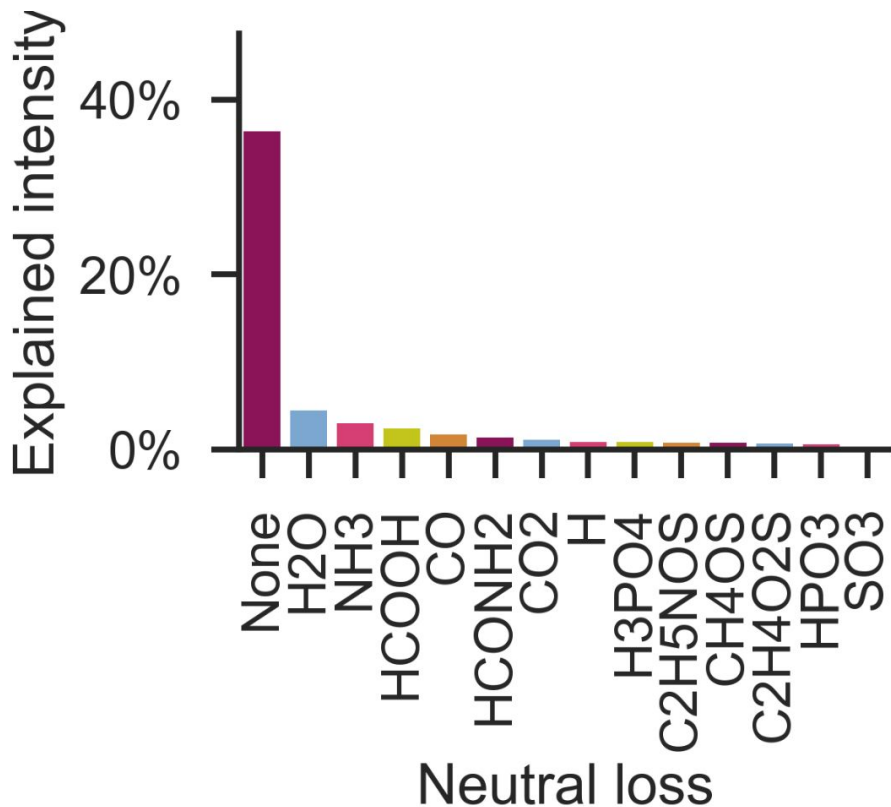
QUIZ: BOTTOM-UP OR IMMUNOPEPTIDOMICS?



QUIZ: BOTTOM-UP OR IMMUNOPEPTIDOMICS?

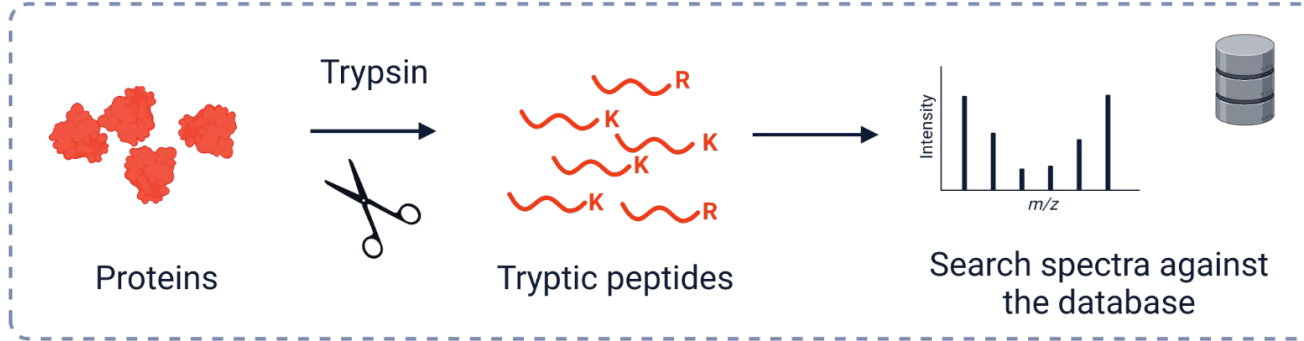


QUIZ: BOTTOM-UP OR IMMUNOPEPTIDOMICS?

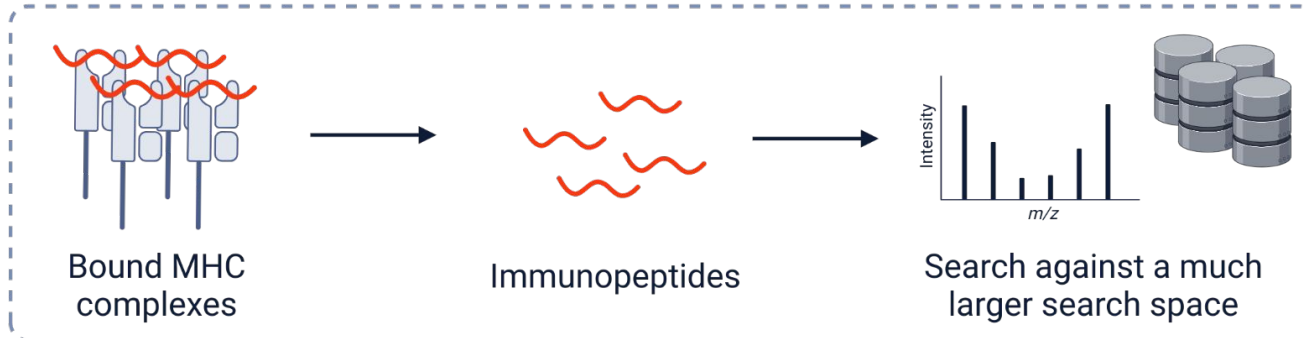


IMMUNOPEPTIDOMICS

Bottom-up Proteomics

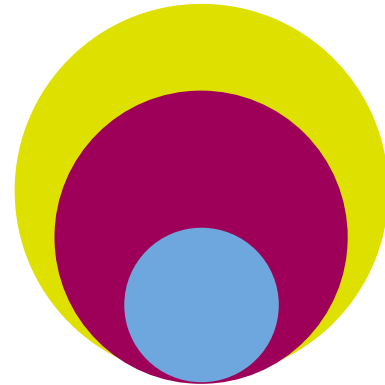


Immunopeptidomics



CHALLENGES IN IMMUNOPEPTIDOMICS ANALYSIS

- Massive search space: all protein subsequences have to be considered
- Non-tryptic peptides
 - Often singly charged: poor ionization and fragmentation coverage
- Present at low abundances

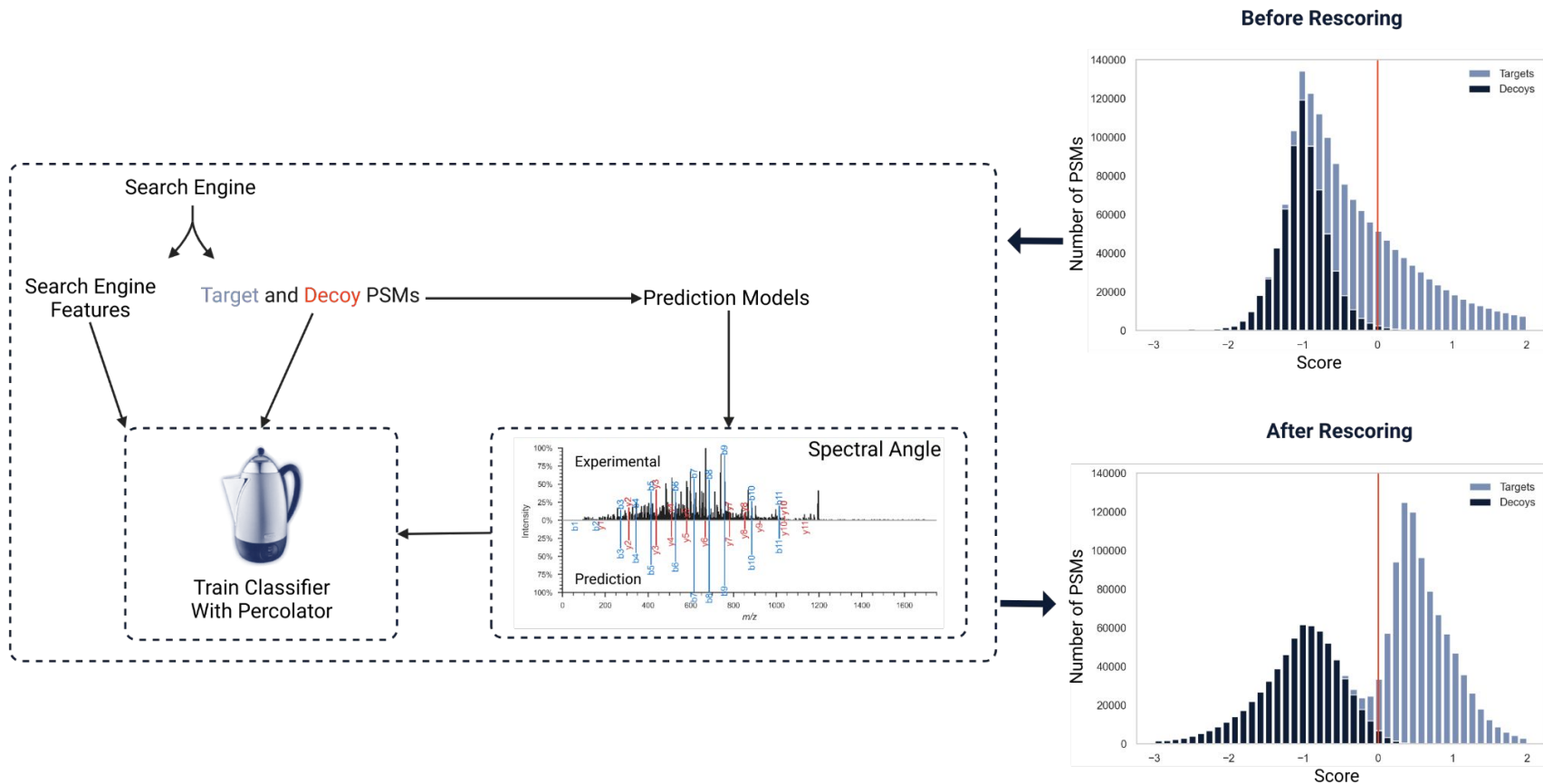


non-tryptic

semi-tryptic

tryptic

PEPTIDE-SPECTRUM-MATCH RESCORING



FEATURE TYPES FOR (IMMUNOPEPTIDE) RESCORING

🌟 Spectral similarity features

- Cosine similarity, spectral contrast angle, spectral entropy, ...

🌟 Retention time features: ΔRT

😞 Ion mobility features

⚠️ Binding affinity features

TOOLS FOR (IMMUNOPEPTIDE) RESCORING

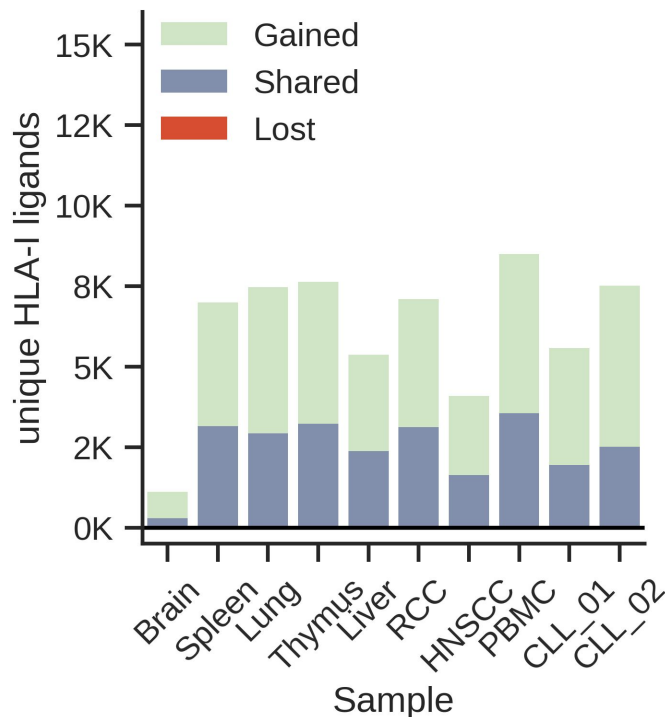
Rescoring pipeline	RT	MS/MS	CCS	BA	Search engine
AlphaPeptDeep	integrated	integrated	integrated	N/A	AlphaPept, MaxQuant, DIA-NN
DeepRescore	AutoRT	pDeep2	N/A	N/A	Comet, MaxQuant, MS-GF+, X!Tandem
INFERYS	N/A	integrated	N/A	N/A	Sequest HT
inSPIRE	Prosit / pyteomics	Prosit / MS ² PIP	N/A	netMHCpan	Mascot, MaxQuant, PEAKS
MHCquant	N/A	N/A	N/A	MHCFlurry	Comet (Integrated)
MS ² Rescore	DeepLC	MS ² PIP	N/A	N/A	MaxQuant, MS-GF+, PEAKS, PeptideShaker, X!Tandem
MSBooster	DIA-NN	DIA-NN	DIA-NN	N/A	MSFragger, MSFragger-DIA
Oktoberfest	Prosit	Prosit	N/A	N/A	MaxQuant, MSFragger, custom

>2-FOLD PEPTIDE INCREASE AFTER RESCORING

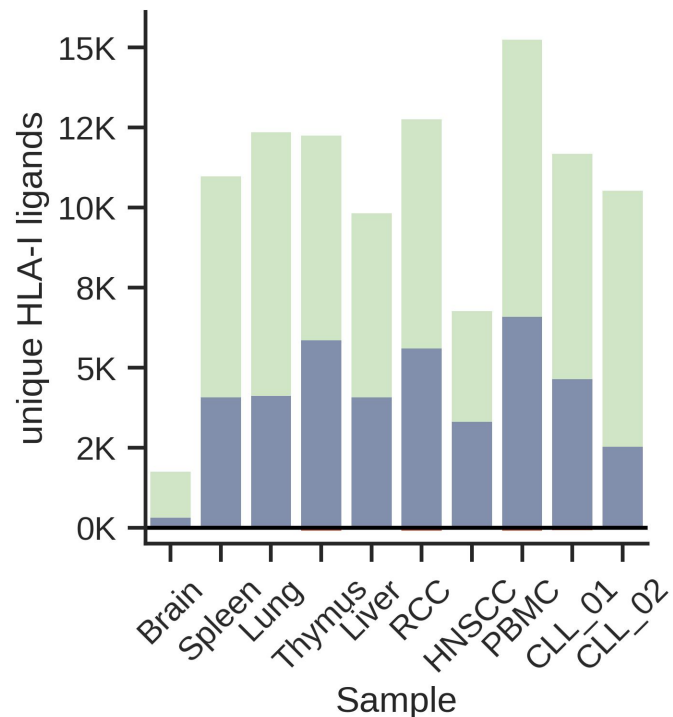


Charlotte Adams

Orbitrap – Prosit 2020 HCD



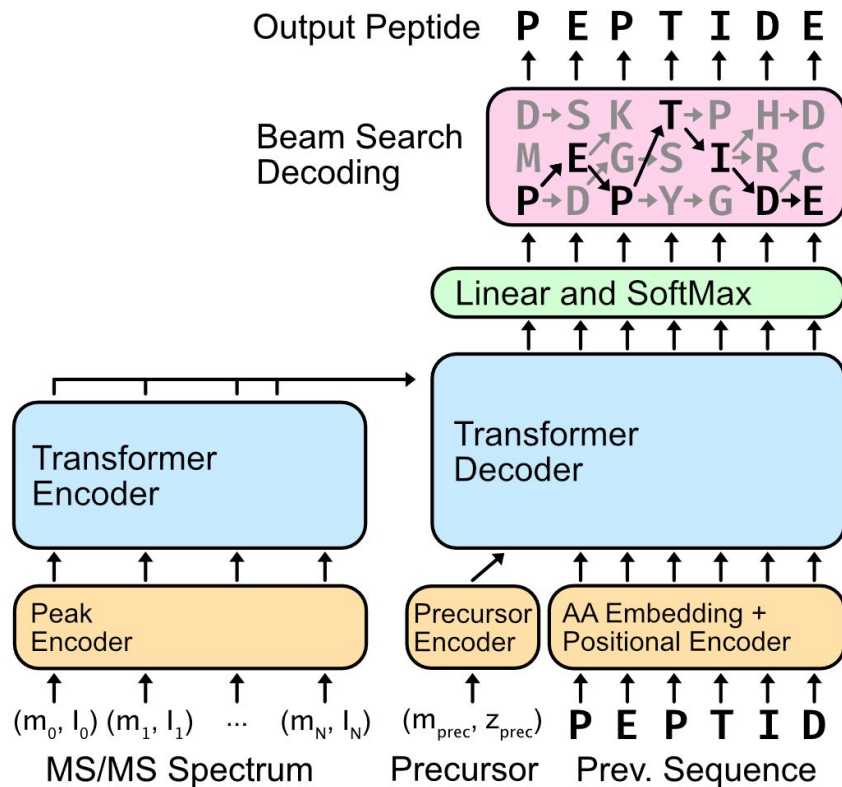
timsTOF – Prosit 2023 TOF



EMERGING SOLUTIONS: *DE NOVO* SEQUENCING

Casanovo is a transformer de novo peptide sequencing tool

- **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

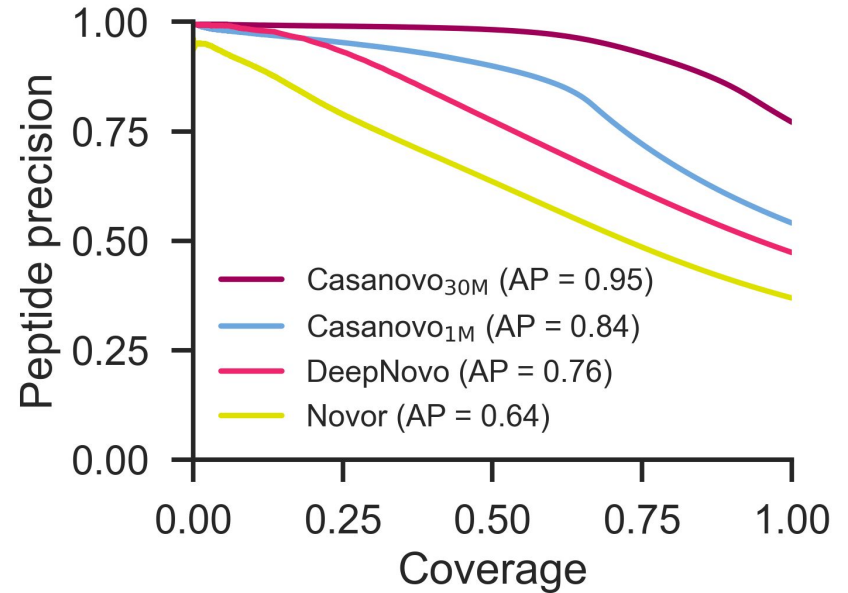


CASANOVO OUTPERFORMS OTHER METHODS



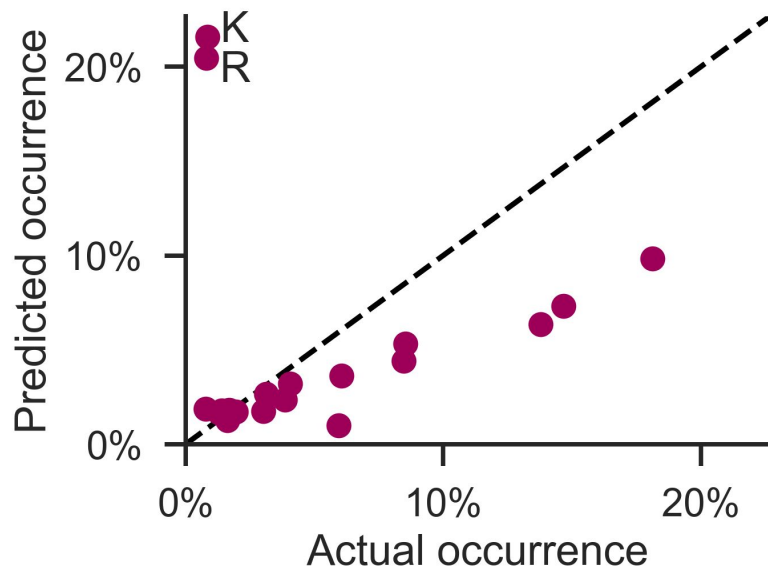
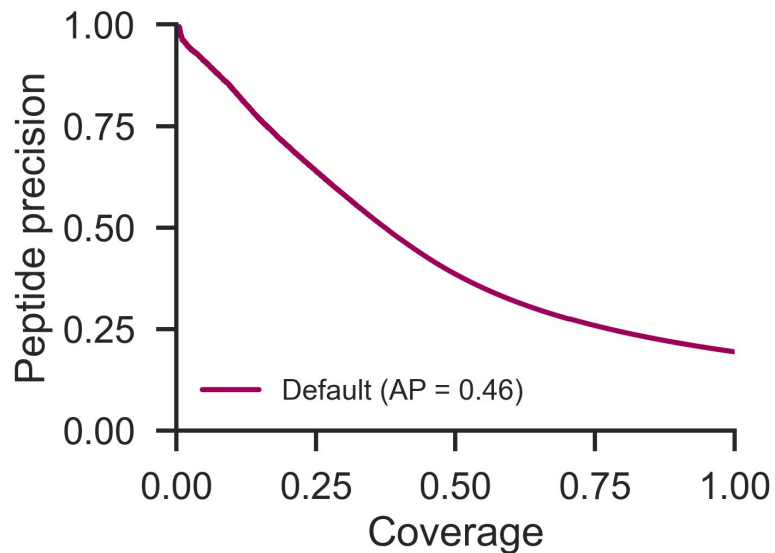
Melih Yilmaz

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



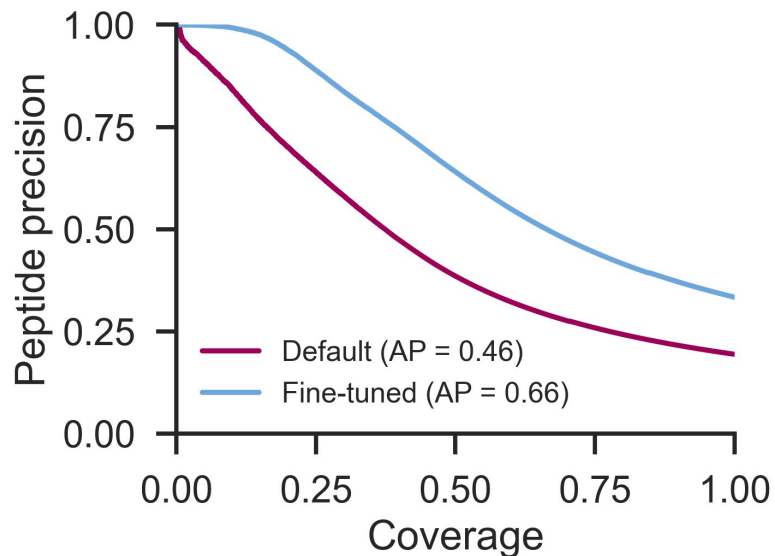
NON-ENZYMATIC FINE-TUNING AVOIDS TRYPTIC BIAS

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

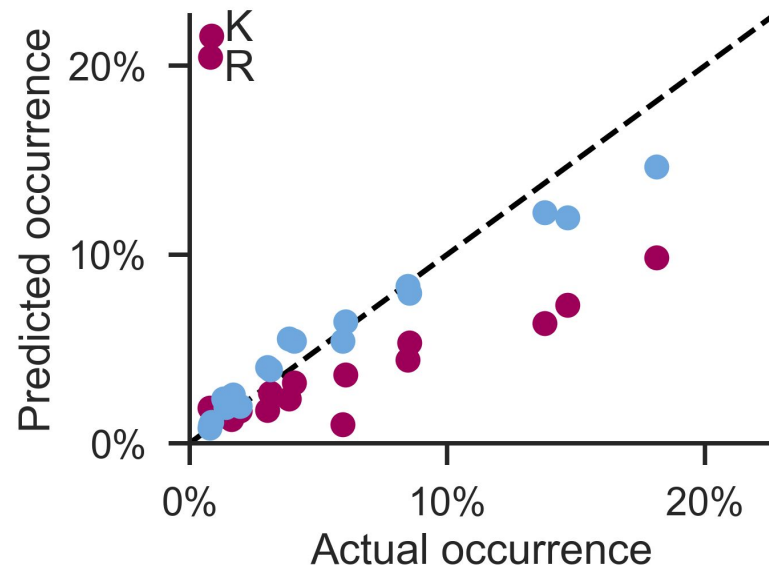


NON-ENZYMATIC FINE-TUNING AVOIDS TRYPTIC BIAS

- 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



TO *DE NOVO* OR NOT TO *DE NOVO*?

Opportunities for *de novo*

- Discovery of novel peptides (neoantigens, viral peptides, ...)
- No need for reference sequence database
- Outperform sequence database searching (?)

Challenges of *de novo*

- No approach for FDR estimation
- Requires higher-quality MS/MS data
- Underperform sequence database searching (?)