

Supplementary Information

Accurate and ultra-fast *de novo* HLA-I immunopeptide sequencing with FoxNovo

Ze-Xuan Chen^{1,2,3}; Chang-Rong You^{1,2,3}; Ching Tarn^{2,3}; Xie-Xuan Zhou³; Wen-Feng Zeng^{2,3,*}

1. Zhejiang University, Hangzhou, China

2. Department of Artificial Intelligence, School of Engineering, Westlake University

3. Center for Infectious Disease Research, School of Medicine, Westlake University

* Email: zengwenfeng@westlake.edu.cn

Supplementary materials

Contents

<u>Supplementary Table 1. Collected HLA-I immunopeptidomics project IDs and numbers of raw files.</u>	2
<u>Supplementary Table 2. FoxNovo identification and ncORF mapping of 42 validated non-canonical peptides.</u>	5
<u>Supplementary Figure 1. Peptide-charge pair distributions before and after weighted sampling.</u>	7
<u>Supplementary Figure 2. Analysis of database search (pFind3) on the peak removal benchmark.</u>	8
<u>Supplementary Figure 3. Determining the QC score cutoff for FoxNovo.</u>	9
<u>Supplementary Figure 4. FoxNovo performance on database search-missed spectra.</u>	10
<u>Supplementary Figure 5. Benchmarking de novo AI models on peak-removal-with-refill test sets.</u>	11

Supplementary Table 1. Collected HLA-I immunopeptidomics project IDs and numbers of raw files.

Project ID	Raw files	Reference
PXD000394	26	Ref ¹
PXD004023	186	Ref ²
PXD005231	41	https://www.ebi.ac.uk/pride/archive/projects/PXD005231
PXD006455	28	Ref ³
PXD006939	52	Ref ⁴
PXD007596	16	Ref ⁵
PXD007679	15	Ref ⁶
PXD007756	68	Ref ⁷
PXD008500	15	Ref ⁸
PXD008937	10	Ref ⁹
PXD011257	1	Ref ¹⁰
PXD011723	30	Ref ¹¹
PXD012348	18	Ref ¹²
PXD012498	6	Ref ¹³
PXD012797	21	Ref ¹⁴
PXD013064	8	Ref ¹⁵
PXD013831	81	Ref ¹⁶
PXD014397	83	Ref ¹⁷
PXD014804	15	Ref ¹⁸
PXD015039	14	Ref ¹⁹
PXD015957	30	Ref ²⁰
PXD018539	23	Ref ²¹
PXD019466	32	https://www.ebi.ac.uk/pride/archive/projects/PXD019466
PXD019643	576	Ref ²²
PXD020186	9	Ref ²²
PXD020750	24	Ref ²³
PXD021177	2	Ref ²⁴
PXD021398	34	https://www.ebi.ac.uk/pride/archive/projects/PXD021398
PXD021755	9	https://www.ebi.ac.uk/pride/archive/projects/PXD021755
PXD022020	7	Ref ²⁵
PXD022150	36	Ref ²⁶
PXD022323	1	https://www.ebi.ac.uk/pride/archive/projects/PXD022323
PXD022884	72	https://www.ebi.ac.uk/pride/archive/projects/PXD022884
PXD022949	7	Ref ²⁷
PXD022950	13	Ref ²⁸
PXD023044	47	Ref ²⁹
PXD023064	7	https://www.ebi.ac.uk/pride/archive/projects/PXD023064
PXD023143	67	Ref ³⁰
PXD023545	65	https://www.ebi.ac.uk/pride/archive/projects/PXD023545
PXD024109	28	Ref ³¹
PXD024562	4	https://www.ebi.ac.uk/pride/archive/projects/PXD024562
PXD025499	73	Ref ³²
PXD025655	11	Ref ³³
PXD026184	20	Ref ³⁴

Project ID	Raw files	Reference
PXD027182	26	https://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX027182
PXD028088	4	https://www.ebi.ac.uk/pride/archive/projects/PXD028088
PXD028633	29	Ref ³⁵
PXD028862	36	Ref ³⁶
PXD028985	62	Ref ³⁷
PXD029567	60	Ref ³⁸
PXD029594	11	https://www.ebi.ac.uk/pride/archive/projects/PXD029594
PXD029613	16	https://www.ebi.ac.uk/pride/archive/projects/PXD029613
PXD029860	3	Ref ³⁹
PXD029882	42	https://www.ebi.ac.uk/pride/archive/projects/PXD029882
PXD029895	3	Ref ⁴⁰
PXD031451	16	https://www.ebi.ac.uk/pride/archive/projects/PXD031451
PXD031709	9	https://www.ebi.ac.uk/pride/archive/projects/PXD031709
PXD033685	3	https://www.ebi.ac.uk/pride/archive/projects/PXD033685
PXD033868	15	Ref ⁴¹
PXD034017	1	Ref ⁴²
PXD034429	13	https://www.ebi.ac.uk/pride/archive/projects/PXD034429
PXD034772	62	Ref ⁴³
PXD034820	6	Ref ⁴⁴
PXD036277	23	https://www.ebi.ac.uk/pride/archive/projects/PXD036277
PXD037365	6	https://www.ebi.ac.uk/pride/archive/projects/PXD037365
PXD037627	8	Ref ⁴⁵
PXD037655	107	Ref ⁴⁶
PXD037837	72	Ref ⁴⁷
PXD040651	196	https://www.ebi.ac.uk/pride/archive/projects/PXD040651
PXD041893	301	Ref ⁴⁸
PXD043057	7	Ref ⁴⁹
PXD043542	16	Ref ⁵⁰
PXD043984	6	https://www.ebi.ac.uk/pride/archive/projects/PXD043984
PXD043989	16	Ref ⁵¹
PXD044085	6	Ref ⁵²
PXD046803	19	Ref ⁵³
PXD046825	9	Ref ⁵⁴
PXD046857	1	Ref ⁵⁵
PXD049996	13	https://www.ebi.ac.uk/pride/archive/projects/PXD049996
PXD051190	1	Ref ⁵⁶
PXD051538	8	Ref ⁵⁷
PXD051734	8	Ref ⁵⁸
PXD052187	9	Ref ⁵⁹
PXD053264	13	Ref ⁶⁰
PXD055042	4	https://www.ebi.ac.uk/pride/archive/projects/PXD055042
PXD055609	5	https://www.ebi.ac.uk/pride/archive/projects/PXD055609
PXD055726	23	Ref ⁶¹
PXD056426	17	Ref ⁶²
PXD059824	63	https://www.ebi.ac.uk/pride/archive/projects/PXD059824
PXD066255	4	https://www.ebi.ac.uk/pride/archive/projects/PXD066255
PXD066752	1	Ref ⁶³

Project ID	Raw files	Reference
PXD003790	17	https://www.ebi.ac.uk/pride/archive/projects/PXD003790
PXD008127	61	Ref ⁶⁴
PXD009738	26	Ref ⁶⁵
PXD009925	4	https://www.ebi.ac.uk/pride/archive/projects/PXD009925
PXD012083	10	Ref ⁶⁶
PXD013649	52	Ref ⁶⁷
PXD014062	8	Ref ⁶⁸
PXD015646	23	Ref ⁶⁹
PXD016060	10	Ref ⁷⁰
PXD016582	14	Ref ⁷¹
PXD020011	96	Ref ⁷²
PXD028364	6	https://www.ebi.ac.uk/pride/archive/projects/PXD028364
PXD031108	11	Ref ⁷³
PXD037270	45	Ref ⁷⁴
PXD037845	9	Ref ⁷⁵
PXD038271	9	Ref ⁷⁶
PXD038371	6	Ref ⁷⁷
PXD044235	18	Ref ⁷⁸
PXD045000	106	Ref ⁷⁹
PXD045265	15	Ref ⁸⁰
PXD046182	33	Ref ⁸¹
PXD046891	2	Ref ⁸²
PXD051490	20	Ref ⁸³
PXD052031	14	Ref ⁸⁴
PXD024533	9	Ref ⁸⁵
PXD027830	9	Ref ⁸⁶
PXD040740	5	http://central.proteomexchange.org/cgi/GetDataset?ID=PXD040740
PXD047303	5	Ref ⁸⁷
PXD065200	2	Ref ⁸⁸
PXD055841	11	https://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PX055841-1
MSV000084442	46	Ref ⁸⁹
MSV000080527	72	Ref ⁹⁰
MSV000084172	339	Ref ⁸⁹

Supplementary Table 2. FoxNovo identification and ncORF mapping of 42 validated non-canonical peptides.

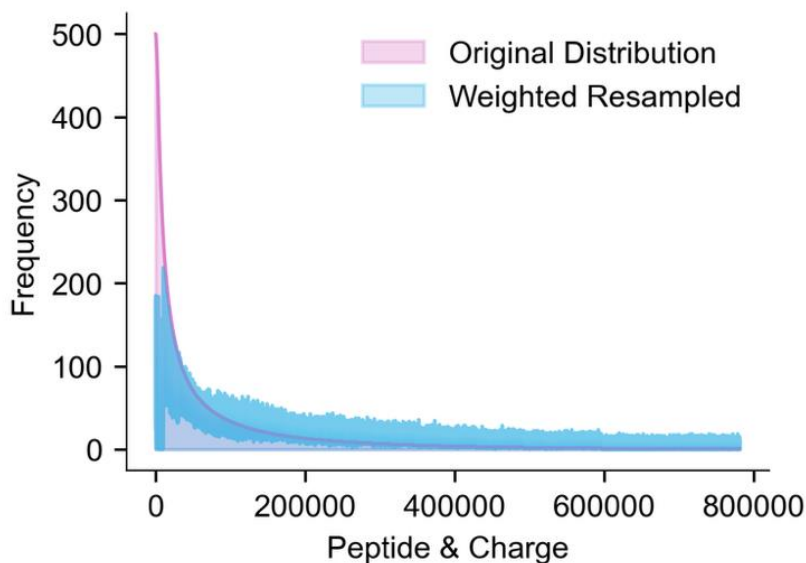
HLA peptide origin	Sequence	Recovered by FoxNovo	Mapped to ncORF database	ncORF match
lncRNA	RFTDVRVQF	Yes	Yes	c8norep34
lncRNA	VTDAQASHIY	Yes	No	N/A
lncRNA	ATEHDKNSTNIY	Yes	Yes	c1riboseqorf140
lncRNA	FIETESQLY	Yes	Yes	c2norep199
lncRNA	RVSHLTSSV	Yes	No	N/A
lncRNA	IFNPVSKNF	Yes	Yes	c2norep207; c2norep212
lncRNA	IWNGKGFHSF	Yes	No	N/A
lncRNA	SYLEPQVTF	Yes	No	N/A
lncRNA	ANRPPTQAL	Yes	No	N/A
lncRNA	RPGPGTRKSL	Yes	Yes	c22riboseqorf33
lncRNA	SYLRRHLDF	Yes	No	N/A
lncRNA	SPREPAGGGQV	Yes	No	N/A
lncRNA	FSDFPGGDKGKFY	Yes	Yes	c1riboseqorf116
lncRNA	SPAQRDRVVM	Yes	Yes	c6riboseqorf130
lncRNA	AAEPDRAVL	Yes	No	N/A
lncRNA	RQFLEPSGF	Yes	Yes	c2riboseqorf65
lncRNA	TPNSLAKPL	Yes	No	N/A
lncRNA	KSDLKPLSY	Yes	No	N/A
lncRNA	NPRNDGIAL	Yes	Yes	c4norep80; c4riboseqorf44; c4riboseqorf46
lncRNA	LPQETRIPL	Yes	No	N/A
lncRNA	DVDIHKDLY	Yes	No	N/A
lncRNA/TEs	SPRSPSPSL	No	No	N/A
lncRNA	RIHDSRSSL	Yes	No	N/A
lncRNA	ASKGAGVSF	Yes	No	N/A
lncRNA	KMEETTVTY	Yes	No	N/A
lncRNA	EYNNIFHTF	Yes	No	N/A
lncRNA	ETDIOMETRY	Yes	Yes	c8riboseqorf79
lncRNA	YLDPAQQNLY	Yes	No	N/A
lncRNA	VPHQPAPPF	Yes	Yes	c20riboseqorf37; c20riboseqorf38
lncRNA	TFQKKTkem	Yes	No	N/A
lncRNA	APKLMVRPF	Yes	Yes	cXriboseqorf15
lncRNA	QTEKENSky	Yes	No	N/A
lncRNA/TEs	KPRHQPGPAL	Yes	No	N/A
lncRNA	APKSSSGFSL	Yes	No	N/A
lncRNA	YSETSLSNGGKY	Yes	No	N/A

lncRNA	RPPHPRPSL	Yes	No	N/A
lncRNA	KFLVQNNVF	Yes	No	N/A
lncRNA	HIKTQDTGY	Yes	No	N/A
lncRNA	VMKGINQGY	Yes	No	N/A
TEs annotated	KVFKNGNAF	Yes	No	N/A
TEs	NMDGPGGHY	Yes	No	N/A
TEs	AAFDRAVHF	Yes	No	N/A

Note:

This table lists the 42 PRM-validated non-canonical melanoma HLA-I peptides from Chong et al⁶⁷. and indicates whether each peptide was recovered by FoxNovo de novo sequencing and whether it mapped to the public ncORF database^{91,92}.

Supplementary Figure 1. Peptide-charge pair distributions before and after weighted sampling.



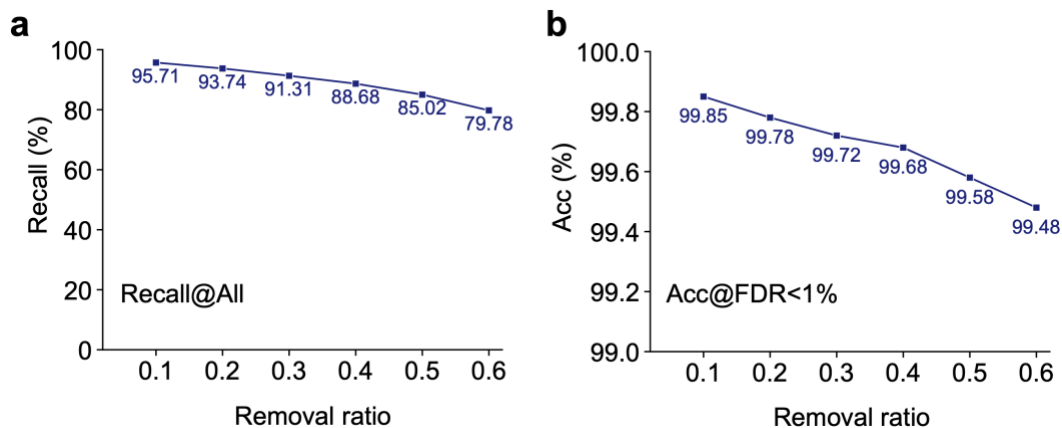
Note:

The training datasets contain 781,588 peptide-charge pairs.

Sampling statistics are shown for one training epoch.

Peptide-charge-weighted sampling reduces the dominance of high-frequency pairs while increasing the sampling frequency of most low-frequency pairs.

Supplementary Figure 2. Analysis of database search (pFind3) on the peak removal benchmark.



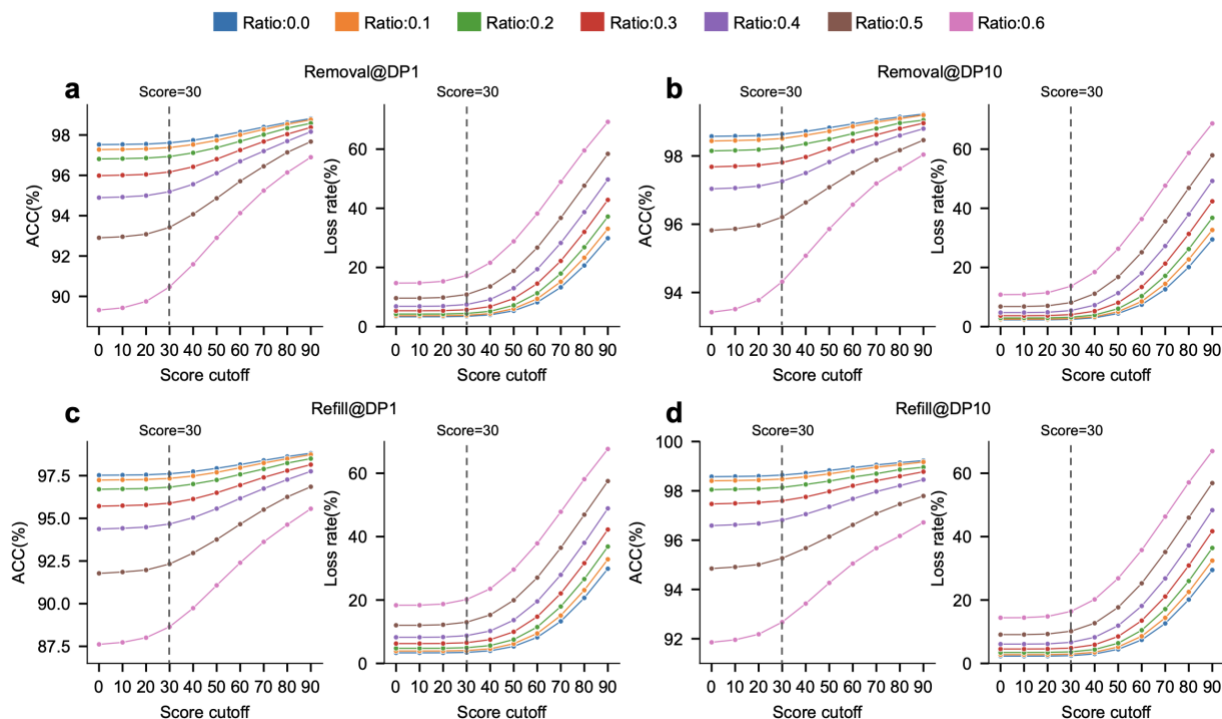
- (a) Recall of pFind3 across all spectra at different peak removal ratios, without applying FDR filtering.
(b) Accuracy of pFind3 at different peak removal levels after applying a <1% PSM-level FDR threshold.

Note:

Database search recall decreased as the peak removal ratio increased, reaching 85.02% and 79.78% at 50% and 60% removal, respectively.

FDR control effectively maintained high identification accuracy, with accuracy remaining above 99% across all peak removal ratios.

Supplementary Figure 3. Determining the QC score cutoff for FoxNovo.



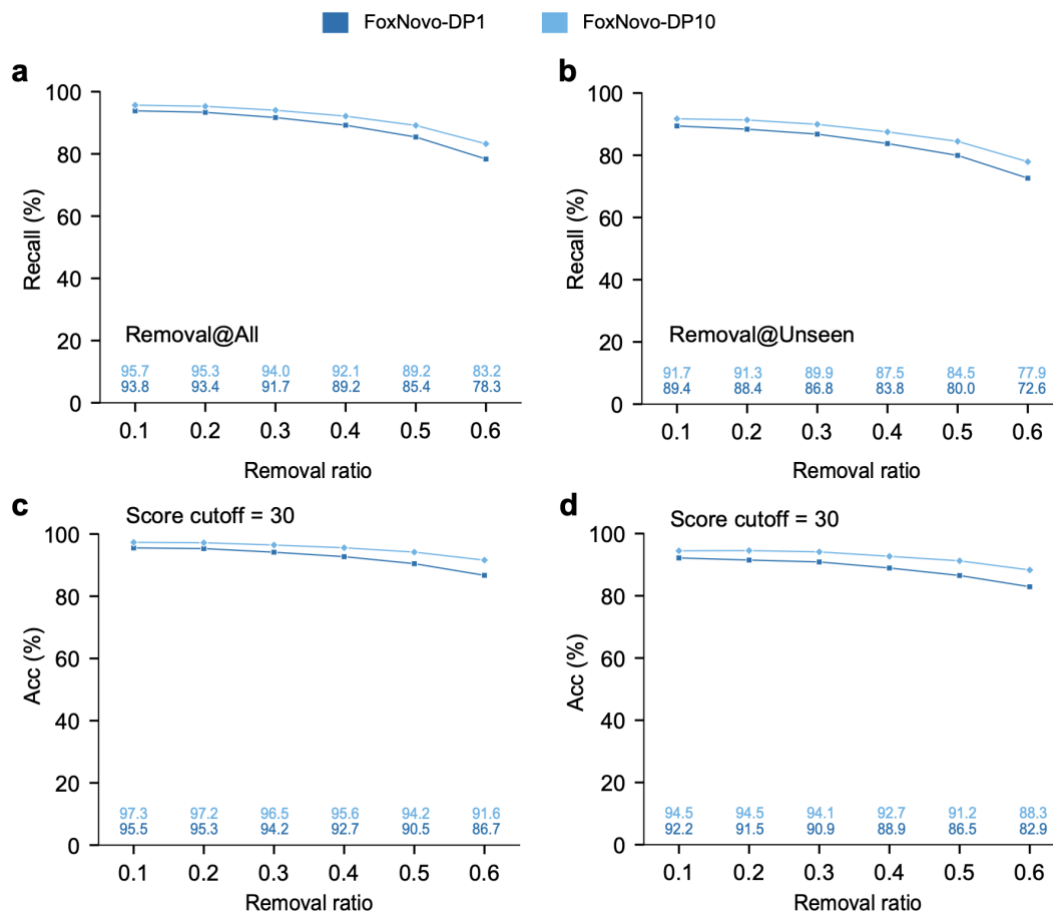
- (a) Accuracy and loss rate at different cutoffs on the peak removal test dataset for FoxNovo-DP1.
- (b) Accuracy and loss rate at different cutoffs on the peak removal test dataset for FoxNovo-DP10.
- (c) Accuracy and loss rate at different cutoffs on the peak-removal-with-refill test dataset for FoxNovo-DP1.
- (d) Accuracy and loss rate at different cutoffs on the peak-removal-with-refill test dataset for FoxNovo-DP10.

Note:

The loss rate is defined as one minus recall after score filtering. The refill process refers to peak-removal-with-refill setting.

The final QC score threshold was set to 30, which provides a good balance between accuracy and loss rate.

Supplementary Figure 4. FoxNovo performance on database search-missed spectra.



(a) Recall of FoxNovo-DP1 and FoxNovo-DP10 on database-search-missed peak-removal spectra across different peak-removal ratios.

(b) Recall of FoxNovo-DP1 and FoxNovo-DP10 on database-search-missed spectra from the unseen-peptide subset across different peak-removal ratios.

(c) Accuracy of FoxNovo-DP1 and FoxNovo-DP10 with QC filtering (score cutoff) on database-search-missed peak-removal spectra across different peak-removal ratios.

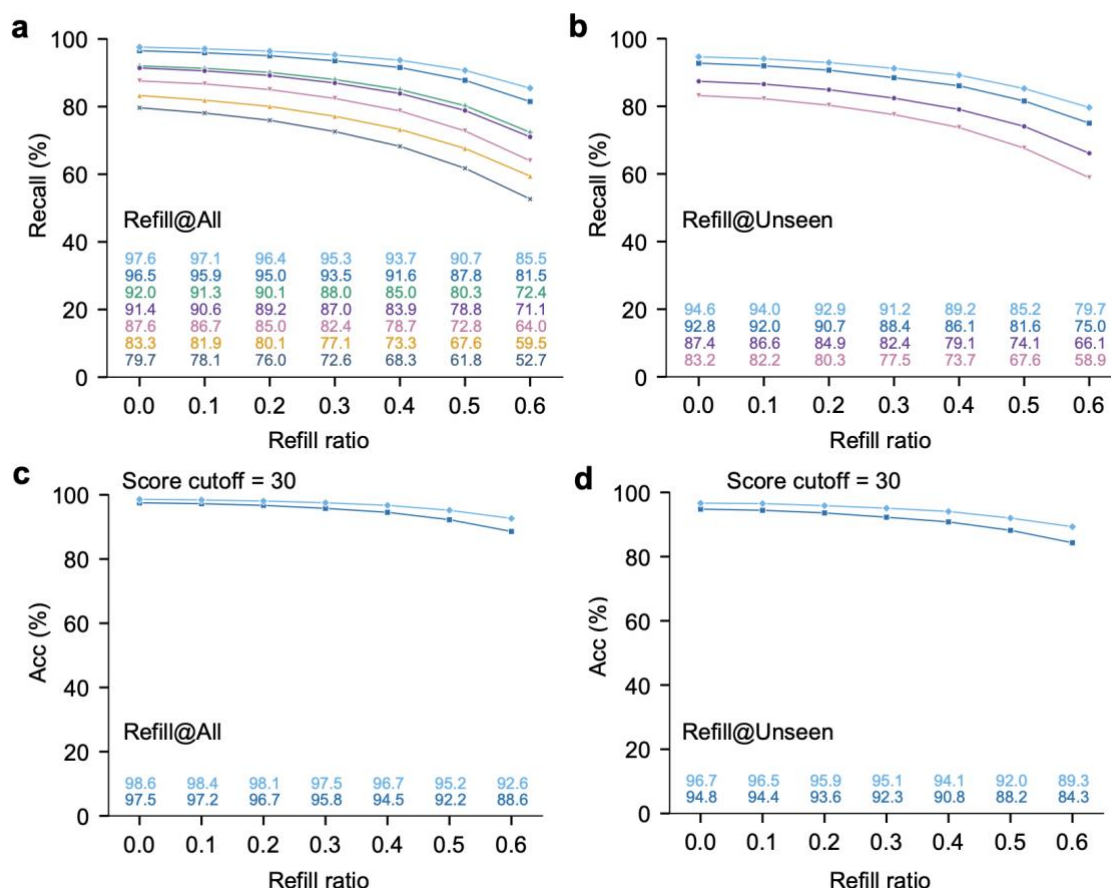
(d) Accuracy of FoxNovo-DP1 and FoxNovo-DP10 with QC filtering on database-search-missed spectra from the unseen-peptide subset across different peak-removal ratios.

Note:

FoxNovo-DP1 achieved recall of 85.4% and 78.3% on database-search-missed perturbed spectra at peak-removal ratios of 50% and 60%, respectively.

Supplementary Figure 5. Benchmarking de novo AI models on peak-removal-with-refill test sets.

FoxNovo-DP10 FoxNovo-DP1 InstaNovo+ Casanovo-HLA-Bm Casanovo-HLA PrimeNovo RNovA



(a) Recall of all de novo models on all spectra in the peak-removal-with-refill benchmark across different refill ratios.

(b) Recall of FoxNovo-DP1 and FoxNovo-DP10 on unseen-peptide spectra in the peak-removal-with-refill benchmark across different refill ratios.

(c) Accuracy of FoxNovo-DP1 and FoxNovo-DP10 with QC filtering (score cutoff) on all spectra in the peak-removal-with-refill benchmark across different refill ratios.

(d) Accuracy of FoxNovo-DP1 and FoxNovo-DP10 with QC filtering on unseen-peptide spectra in the peak-removal-with-refill benchmark across different refill ratios.

Note:

FoxNovo-DP1 achieved recall of 87.8% and 81.5% on perturbed spectra with refill ratios of 50% and 60%, respectively.

References

1. Bassani-Sternberg, M., Platscher-Frankild, S., Jensen, L. J. & Mann, M. 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 (2015).
2. Pearson, H. *et al.* MHC class I-associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* **126**, 4690–4701 (2016).
3. Kaur, G. *et al.* Structural and regulatory diversity shape HLA-C protein expression levels. *Nat. Commun.* **8**, 15924 (2017).
4. Chong, C. *et al.* High-throughput and Sensitive Immunopeptidomics Platform Reveals Profound Interferon-Mediated Remodeling of the Human Leukocyte Antigen (HLA) Ligandome. *Mol. Cell. Proteomics* **17**, 533–548 (2018).
5. Komov, L. *et al.* Cell Surface MHC Class I Expression Is Limited by the Availability of Peptide-Receptive “Empty” Molecules Rather than by the Supply of Peptide Ligands. *Proteomics* **18**, 1700248 (2018).
6. Sanz-Bravo, A. *et al.* Allele-specific Alterations in the Peptidome Underlie the Joint Association of HLA-A*29:02 and Endoplasmic Reticulum Aminopeptidase 2 (ERAP2) with Birdshot Chorioretinopathy. *Mol. Cell. Proteomics* **17**, 1564–1577 (2018).
7. Yair-Sabag, S. *et al.* The Peptide Repertoire of HLA-B27 may include Ligands with Lysine at P2 Anchor Position. *Proteomics* **18**, 1700249 (2018).
8. Sanz-Bravo, A. *et al.* Ranking the Contribution of Ankylosing Spondylitis-associated Endoplasmic Reticulum Aminopeptidase 1 (ERAP1) Polymorphisms to Shaping the HLA-B*27 Peptidome. *Mol. Cell. Proteomics* **17**, 1308–1323 (2018).
9. Zeiner, P. S. *et al.* CD74 regulates complexity of tumor cell HLA class II peptidome in brain metastasis and is a positive prognostic marker for patient survival. *Acta Neuropathol. Commun.* **6**, 18 (2018).
10. Demmers, L. C., Heck, A. J. R. & Wu, W. Pre-fractionation Extends but also Creates a Bias in the Detectable HLA Class I Ligandome. *J. Proteome Res.* **18**, 1634–1643 (2019).
11. Andreatta, M. *et al.* MS-Rescue: A Computational Pipeline to Increase the Quality and Yield of Immunopeptidomics Experiments. *Proteomics* **19**, 1800357 (2019).
12. Guasp, P. *et al.* Redundancy and Complementarity between ERAP1 and ERAP2 Revealed by their Effects on the Behcet’s Disease-associated HLA-B*51 Peptidome. *Mol. Cell. Proteomics* **18**, 1491–1510 (2019).

13. Sturm, T. *et al.* Mild Acid Elution and MHC Immunoaffinity Chromatography Reveal Similar Albeit Not Identical Profiles of the HLA Class I Immuno-peptidome. *J. Proteome Res.* **20**, 289–304 (2021).
14. Ghosh, M. *et al.* Guidance Document: Validation of a High-Performance Liquid Chromatography-Tandem Mass Spectrometry Immuno-peptidomics Assay for the Identification of HLA Class I Ligands Suitable for Pharmaceutical Therapies. *Mol. Cell. Proteomics* **19**, 432–443 (2020).
15. Chen, L. *et al.* Identification of an Unconventional Subpeptidome Bound to the Behçet's Disease-associated HLA-B*51:01 that is Regulated by Endoplasmic Reticulum Amino-peptidase 1 (ERAP1). *Mol. Cell. Proteomics* **19**, 871–883 (2020).
16. Solleder, M. *et al.* Mass Spectrometry Based Immuno-peptidomics Leads to Robust Predictions of Phosphorylated HLA Class I Ligands. *Mol. Cell. Proteomics* **19**, 390–404 (2020).
17. Faridi, P. *et al.* Spliced Peptides and Cytokine-Driven Changes in the Immuno-peptidome of Melanoma. *Cancer Immunol. Res.* **8**, 1322–1334 (2020).
18. Lorente, E. *et al.* Substantial Influence of ERAP2 on the HLA-B*40:02 Peptidome: Implications for HLA-B*27-Negative Ankylosing Spondylitis. *Mol. Cell. Proteomics* **18**, 2298–2309 (2019).
19. Pandey, K. *et al.* In-depth mining of the immuno-peptidome of an acute myeloid leukemia cell line using complementary ligand enrichment and data acquisition strategies. *Mol. Immunol.* **123**, 7–17 (2020).
20. Kalaora, S. *et al.* Immunoproteasome expression is associated with better prognosis and response to checkpoint therapies in melanoma. *Nat. Commun.* **11**, 896 (2020).
21. Ehx, G. *et al.* Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. *Immunity* **54**, 737-752.e10 (2021).
22. Marcu, A. *et al.* HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J. Immunother. Cancer* **9**, e002071 (2021).
23. Olsson, N. *et al.* An Integrated Genomic, Proteomic, and Immuno-peptidomic Approach to Discover Treatment-Induced Neoantigens. *Front. Immunol.* **12**, 662443 (2021).
24. Bauzá-Martínez, J., Heck, A. J. R. & Wu, W. HLA-B and cysteinylated ligands distinguish the antigen presentation landscape of extracellular vesicles. *Commun. Biol.* **4**, 825 (2021).
25. Martin, A. D. *et al.* Re-examination of MAGE-A3 as a T-cell Therapeutic Target. *J. Immunother.* **44**, 95–105 (2021).

26. Kalaora, S. *et al.* Identification of bacteria-derived HLA-bound peptides in melanoma. *Nature* **592**, 138–143 (2021).
27. Qi, Y. A. *et al.* Proteogenomic Analysis Unveils the HLA Class I-Presented Immunoepitidome in Melanoma and EGFR-Mutant Lung Adenocarcinoma. *Mol. Cell. Proteomics* **20**, 100136 (2021).
28. Pak, H. *et al.* Sensitive Immunoepitidomics by Leveraging Available Large-Scale Multi-HLA Spectral Libraries, Data-Independent Acquisition, and MS/MS Prediction. *Mol. Cell. Proteomics* **20**, 100080 (2021).
29. Goncalves, G. *et al.* IFN γ Modulates the Immunoepitidome of Triple Negative Breast Cancer Cells by Enhancing and Diversifying Antigen Processing and Presentation. *Front. Immunol.* **12**, 645770 (2021).
30. de Beijer, M. T. A. *et al.* Immunoepitidome of hepatocytes isolated from patients with HBV infection and hepatocellular carcinoma. *JHEP Reports* **4**, 100576 (2022).
31. Venema, W. J. *et al.* ERAP2 Increases the Abundance of a Peptide Submotif Highly Selective for the Birdshot Uveitis-Associated HLA-A29. *Front. Immunol.* **12**, 634441 (2021).
32. Nagler, A. *et al.* Identification of presented SARS-CoV-2 HLA class I and HLA class II peptides using HLA peptidomics. *Cell Rep.* **35**, 109305 (2021).
33. Parker, R. *et al.* The Choice of Search Engine Affects Sequencing Depth and HLA Class I Allele-Specific Peptide Repertoires. *Mol. Cell. Proteomics* **20**, 100124 (2021).
34. Sudhir, P.-R., Lin, T.-D. & Zhang, Q. HLA Allele-Specific Quantitative Profiling of Type 1 Diabetic B Lymphocyte Immunoepitidome. *J. Proteome Res.* **21**, 250–264 (2022).
35. Kovalchik, K. A. *et al.* MhcVizPipe: A Quality Control Software for Rapid Assessment of Small- to Large-Scale Immunoepitidome Datasets. *Mol. Cell. Proteomics* **21**, 100178 (2022).
36. Meyer, S. *et al.* Prevalent and immunodominant CD8 T cell epitopes are conserved in SARS-CoV-2 variants. *Cell Rep.* **42**, 111995 (2023).
37. Habel, J. R. *et al.* HLA-A*11:01-restricted CD8⁺ T cell immunity against influenza A and influenza B viruses in Indigenous and non-Indigenous people. *PLoS Pathog.* **18**, e1010337 (2022).
38. Nelde, A. *et al.* Increased soluble HLA in COVID-19 present a disease-related, diverse immunoepitidome associated with T cell immunity. *iScience* **25**, 105643 (2022).
39. Stopfer, L. E. *et al.* MEK inhibition enhances presentation of targetable MHC-I tumor antigens in mutant melanomas. *Proc. Natl Acad. Sci. USA* **119**, 2208900119 (2022).

40. Temponeras, I. *et al.* ERAP2 Inhibition Induces Cell-Surface Presentation by MOLT-4 Leukemia Cancer Cells of Many Novel and Potentially Antigenic Peptides. *Int. J. Mol. Sci.* **23**, 1913 (2022).
41. Declercq, A. *et al.* MS2Rescore: Data-Driven Rescoring Dramatically Boosts Immunopeptide Identification Rates. *Mol. Cell. Proteomics* **21**, 100266 (2022).
42. Bernhardt, M. *et al.* Extending the Mass Spectrometry-Detectable Landscape of MHC Peptides by Use of Restricted Access Material. *Anal. Chem.* **94**, 14214–14222 (2022).
43. Kraemer, A. I. *et al.* The immunopeptidome landscape associated with T cell infiltration, inflammation and immune editing in lung cancer. *Nat. Cancer* **4**, 608–628 (2023).
44. Merlotti, A. *et al.* Noncanonical splicing junctions between exons and transposable elements represent a source of immunogenic recurrent neo-antigens in patients with lung cancer. *Sci. Immunol.* **8**, eabm6359 (2023).
45. Sherwood, M. *et al.* Identification of natural Zika virus peptides presented on the surface of paediatric brain tumour cells by HLA class I. *PLoS One* **20**, e0335726 (2025).
46. Tretter, C. *et al.* Proteogenomic analysis reveals RNA as a source for tumor-agnostic neoantigen identification. *Nat. Commun.* **14**, 4632 (2023).
47. Leddy, O. K., White, F. M. & Bryson, B. D. Leveraging Immunopeptidomics To Study and Combat Infectious Disease. *mSystems* **6**, e00310-21 (2021).
48. Champagne, J. *et al.* Adoptive T cell therapy targeting an inducible and broadly shared product of aberrant mRNA translation. *Immunity* **58**, 247-262.e9 (2025).
49. Apavaloaei, A. *et al.* Tumor antigens preferentially derive from unmutated genomic sequences in melanoma and non-small cell lung cancer. *Nat. Cancer* **6**, 1419–1437 (2025).
50. Salek, M. *et al.* optiPRM: A Targeted Immunopeptidomics LC-MS Workflow With Ultra-High Sensitivity for the Detection of Mutation-Derived Tumor Neoepitopes From Limited Input Material. *Mol. Cell. Proteomics* **23**, 100825 (2024).
51. Ferreira, H. J. *et al.* Immunopeptidomics-based identification of naturally presented non-canonical circRNA-derived peptides. *Nat. Commun.* **15**, 2357 (2024).
52. Kochavi, A. *et al.* Chemotherapeutic agents and leucine deprivation induce codon-biased aberrant protein production in cancer. *Nucleic Acids Res.* **52**, 13964–13979 (2024).
53. Blanchard, T. *et al.* LOXHD1 is an oncofusion-regulated antigen of ewing sarcoma. *Sci. Rep.* **15**, 13007 (2025).

54. Kaur, A. *et al.* Mass Spectrometric Profiling of HLA-B44 Peptidomes Provides Evidence for Tapasin-Mediated Tryptophan Editing. *The Journal of Immunology* **211**, 1298–1307 (2023).
55. Yu, K. K. H. *et al.* Investigative needle core biopsies support multimodal deep-data generation in glioblastoma. *Nat. Commun.* **16**, 3957 (2025).
56. Gerke, C. *et al.* Multimodal HLA-I genotype regulation by human cytomegalovirus US10 and resulting surface patterning. *Elife* **13**, e85560 (2024).
57. Wang, Y. *et al.* Identification of HLA-A*11:01 and A*02:01-Restricted EBV Peptides Using HLA Peptidomics. *Viruses* **16**, 669 (2024).
58. Füchsl, F. *et al.* High-resolution profile of neoantigen-specific TCR activation links moderate stimulation to increased resilience of engineered TCR-T cells. *Nat. Commun.* **15**, 10520 (2024).
59. Caron, E. *et al.* Integrating Machine Learning-Enhanced Immunopeptidomics and SARS-CoV-2 Population-Scale Analyses Unveils Novel Antigenic Features for Next-Generation COVID-19 Vaccines. Preprint at <https://doi.org/10.21203/rs.3.rs-3914861/v1> (2024).
60. Blunt, M. D. *et al.* The nuclear export protein XPO1 provides a peptide ligand for natural killer cells. *Sci. Adv.* **10**, eado6566 (2024).
61. Leddy, O., Yuki, Y., Carrington, M., Bryson, B. D. & White, F. M. PathMHC: a workflow to selectively target pathogen-derived MHC peptides in discovery immunopeptidomics experiments for vaccine target identification. Preprint at <https://doi.org/10.1101/2024.09.11.612454> (2024).
62. Shapiro, I. E. *et al.* Deleterious KOs in the HLA Class I Antigen Processing and Presentation Machinery Induce Distinct Changes in the Immunopeptidome. *Mol. Cell. Proteomics* **24**, 100951 (2025).
63. Nikopaschou, M. *et al.* ERAP1 Allotypes 2 and 10 Differentially Regulate the Immunopeptidome of Melanocytes. *Immunology* **177**, 398–412 (2026).
64. Shraibman, B. *et al.* Identification of Tumor Antigens Among the HLA Peptidomes of Glioblastoma Tumors and Plasma. *Mol. Cell. Proteomics* **18**, 1255–1268 (2019).
65. Ternette, N. *et al.* Immunopeptidomic Profiling of HLA-A2-Positive Triple Negative Breast Cancer Identifies Potential Immunotherapy Target Antigens. *Proteomics* **18**, 1700465 (2018).
66. Narayan, R. *et al.* Acute myeloid leukemia immunopeptidome reveals HLA presentation of mutated nucleophosmin. *PLoS One* **14**, e0219547 (2019).

67. Chong, C. *et al.* Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat. Commun.* **11**, 1293 (2020).
68. Zhao, Q. *et al.* Proteogenomics Uncovers a Vast Repertoire of Shared Tumor-Specific Antigens in Ovarian Cancer. *Cancer Immunol. Res.* **8**, 544–555 (2020).
69. Bettencourt, P. *et al.* Identification of antigens presented by MHC for vaccines against tuberculosis. *NPJ Vaccines* **5**, 2 (2020).
70. Rijensky, N. M. *et al.* Identification of Tumor Antigens in the HLA Peptidome of Patient-derived Xenograft Tumors in Mouse. *Mol. Cell. Proteomics* **19**, 1360–1374 (2020).
71. Demmers, L. C. *et al.* Single-cell derived tumor organoids display diversity in HLA class I peptide presentation. *Nat. Commun.* **11**, 5338 (2020).
72. Marino, F. *et al.* Biogenesis of HLA Ligand Presentation in Immune Cells Upon Activation Reveals Changes in Peptide Length Preference. *Front. Immunol.* **11**, 1981 (2020).
73. Nicholas, B. *et al.* Identification of neoantigens in oesophageal adenocarcinoma. *Immunology* **168**, 420–431 (2023).
74. Srivastava, M. *et al.* Proteogenomic identification of Hepatitis B virus (HBV) genotype-specific HLA-I restricted peptides from HBV-positive patient liver tissues. *Front. Immunol.* **13**, 1032716 (2022).
75. Pandey, K. *et al.* Multiple Classes of Antigen Contribute to the Antigenic Landscape of Mesothelioma. *Mol. Cell. Proteomics* **24**, 100925 (2025).
76. Pandey, K. *et al.* A combined immunopeptidomics, proteomics, and cell surface proteomics approach to identify immunotherapy targets for diffuse intrinsic pontine glioma. *Front. Oncol.* **13**, 1192448 (2023).
77. Wang, S. S. *et al.* Endogenous H3.3K27M derived peptide restricted to HLA-A*02:01 is insufficient for immune-targeting in diffuse midline glioma. *Mol. Ther. Oncolytics* **30**, 167–180 (2023).
78. Chen, J. *et al.* Antiviral CD8⁺ T-cell immune responses are impaired by cigarette smoke and in COPD. *European Respiratory Journal* **62**, 2201374 (2023).
79. Menon, T. *et al.* CD8⁺ T-cell responses towards conserved influenza B virus epitopes across anatomical sites and age. *Nat. Commun.* **15**, 3387 (2024).
80. Carré, A. *et al.* Interferon- α promotes HLA-B-restricted presentation of conventional and alternative antigens in human pancreatic β -cells. *Nat. Commun.* **16**, 765 (2025).

81. Liao, H. *et al.* MARS an improved de novo peptide candidate selection method for non-canonical antigen target discovery in cancer. *Nat. Commun.* **15**, 661 (2024).
82. Wohlfarth, J. *et al.* Loss of p14 diminishes immunogenicity in melanoma via non-canonical Wnt signaling by reducing the peptide surface density. *Mol. Oncol.* **18**, 2449–2470 (2024).
83. Kessler, A. L. *et al.* HLA I immunopeptidome of synthetic long peptide pulsed human dendritic cells for therapeutic vaccine design. *NPJ Vaccines* **10**, 12 (2025).
84. Tsao, H.-W. *et al.* Targeting the aminopeptidase ERAP enhances antitumor immunity by disrupting the NKG2A-HLA-E inhibitory checkpoint. *Immunity* **57**, 2863–2878.e12 (2024).
85. Kikuchi, Y. *et al.* CD8⁺ T-cell Immune Surveillance against a Tumor Antigen Encoded by the Oncogenic Long Noncoding RNA *PVT1*. *Cancer Immunol. Res.* **9**, 1342–1353 (2021).
86. Watts, E. *et al.* Characterization of HLA-A*02:01 MHC Immunopeptide Antigens Enhanced by Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **93**, 13134–13142 (2021).
87. Tokita, S. *et al.* Identification of immunogenic HLA class I and II neoantigens using surrogate immunopeptidomes. *Sci. Adv.* **10**, eado6491 (2024).
88. Goh, S. J. R. *et al.* HLA-A*02:01 Presents Penicillin-Modified Cysteinylated Peptides for T Cell Recognition. *Allergy* **80**, 3165–3177 (2025).
89. Sarkizova, S. *et al.* A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
90. Abelin, J. G. *et al.* Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-allelic Cells Enables More Accurate Epitope Prediction. *Immunity* **46**, 315–326 (2017).
91. Desiere, F. The PeptideAtlas project. *Nucleic Acids Res.* **34**, D655–D658 (2006).
92. Mudge, J. M. *et al.* Standardized annotation of translated open reading frames. *Nat. Biotechnol.* **40**, 994–999 (2022).