Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

☐ n/a - Confirmed
☐ ☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐ ☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐ ☐ The statistical test(s) used AND whether they are one- or two-sided
☐ ☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☐ ☐ A description of all covariates tested
☐ ☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐ ☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐ ☐ For null hypothesis testing, the test statistic (e.g. F, t, χ²) with confidence intervals, effect sizes, degrees of freedom and P value noted. Give P values as exact values whenever suitable.
☐ ☐ For Bayesian analyses, information on the choice of priors and Markov chain Monte Carlo settings
☐ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐ ☐ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Software and code

Policy information about availability of computer code

Data collection: N/A. No data collection involved.

Data analysis: AlphaPept v0.3.31 for RAW data extraction, pFind v3.1.5 for open-search. DIA-Umpire combined with MSFragger v3.4 for direct-DIA search. DIA-NN v1.8.1 for DIA spectral library search. MixMHCpred v2.1 for HLA peptide deconvolution. BertViz v1.4.0 for transformers’ attention visualization. LogoMaker v0.8 for logo plots. AlphaViz on pepdeep_latest branch (https://github.com/MannLabs/alphaviz/tree/peptidepdeep_latest) for MS2 and elution plots.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. Github). See the Nature Portfolio guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, wherever applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The reviewed protein sequence databases are downloaded from uniprot: https://www.uniprot.org/proteomes/UP000005640 for human, https://www.uniprot.org/proteomes/UP000000625 for E. coli, https://www.uniprot.org/proteomes/UP000001744 for fission yeast, and https://www.uniprot.org/proteomes/UP00000803 for drosophila.

The training and testing data were from ProteomeXchange with accession codes: PXD010595 [https://www.ebi.ac.uk/pride/archive/projects/PXD010595], PXD004732 [https://www.ebi.ac.uk/pride/archive/projects/PXD004732], PXD021013 [https://www.ebi.ac.uk/pride/archive/projects/PXD021013], PXD009449 [https://www.ebi.ac.uk/pride/archive/projects/PXD009449], PXD000138 [https://www.ebi.ac.uk/pride/archive/projects/PXD000138], PXD019854 [https://www.ebi.ac.uk/pride/archive/projects/PXD019854], PXD019086 [https://www.ebi.ac.uk/pride/archive/projects/PXD019086], PXD004452 [https://www.ebi.ac.uk/pride/archive/projects/PXD004452], PXD014525 [https://www.ebi.ac.uk/pride/archive/projects/PXD014525], PXD017476 [https://www.ebi.ac.uk/pride/archive/projects/PXD017476], PXD019347 [https://www.ebi.ac.uk/pride/archive/projects/PXD019347], PXD021318 [https://www.ebi.ac.uk/pride/archive/projects/PXD021318], PXD026805 [https://www.ebi.ac.uk/pride/archive/projects/PXD026805], PXD026824 [https://www.ebi.ac.uk/pride/archive/projects/PXD026824], PXD029545 [https://www.ebi.ac.uk/pride/archive/projects/PXD029545], PXD000269 [https://www.ebi.ac.uk/pride/archive/projects/PXD000269], and PXD001250 [https://www.ebi.ac.uk/pride/archive/projects/PXD001250]. The mono-allelic HLA DDA dataset was downloaded from MassiVE with accession code MSV000084172 [https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=77080ee96db0484d9cf6abf5856b5b66].

The tumor HLA dataset was downloaded from ProteomeXchange with accession code PXD004894 [https://www.ebi.ac.uk/pride/archive/projects/PXD004894]. HLA DIA data and the MaxQuant results of DDA data from the RA957 cell line were downloaded from PRIDE with accession code PXD002950 [https://www.ebi.ac.uk/pride/archive/projects/PXD002950]. HLA DIA results of PEAKS-Online were downloaded from https://www.nature.com/articles/s41467-022-30867-7#Sec17. Only DIA data and the MaxQuant results of DDA data from the RA957 cell line were downloaded from PRIDE with accession code PXD022954 [https://www.ebi.ac.uk/pride/archive/projects/PXD022954].

For different collisional energies, and different PTMs. 21 PTMs are generated by ProteomeTools project from Kuster lab.

Overall, nearly 40 million spectra were used to train and test the DL models. There samples were chosen to cover different instruments, different collisional energies, and different PTMs. 21 PTMs are generated by ProteomeTools project from Kuster lab.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender

Use the terms sex (biological attribute) and gender (shaped by social and cultural circumstances) carefully in order to avoid confusing both terms. Indicate if findings apply to only one sex or gender; describe whether sex and gender were considered in study design whether sex and/or gender was determined based on self-reporting or assigned and methods used. Provide in the source data disaggregated sex and gender data where this information has been collected, and consent has been obtained for sharing of individual-level data; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex- and gender-based analyses where performed, justify reasons for lack of sex- and gender-based analysis.

Population characteristics

Describe the covariate-relevant population characteristics of the human research participants (e.g. age, genotypic information, past and current diagnosis and treatment categories). If you filled out the behavioural & social sciences study design questions and have nothing to add here, write “See above.”

Recruitment

Describe how participants were recruited. Outline any potential self-selection bias or other biases that may be present and how these are likely to impact results.

Ethics oversight

Identify the organization(s) that approved the study protocol.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

Overall, nearly 40 million spectra were used to train and test the DL models. There samples were chosen to cover different instruments, different collisional energies, and different PTMs. 21 PTMs are generated by ProteomeTools project from Kuster lab.
Data exclusions
No data exclusions are involved.

Replication
No replication is involved.

Randomization
No randomization is involved.

Blinding
No blinding is involved as we did not use clinical samples and no biological conclusions were made.

Reporting for specific materials, systems and methods
We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems
n/a
- Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology and archaeology
- Animals and other organisms
- Clinical data
- Dual use research of concern

Methods
n/a
- Involved in the study
- ChiP-seq
- Flow cytometry
- MRI-based neuroimaging