Article

ImmuneApp for HLA-I epitope prediction and immunopeptidome analysis

Haodong Xu [1,2,7] ✉, Ruifeng Hu [2,3,4,7], Xianjun Dong [3,4], Lan Kuang[1], Wenchao Zhang[1], Chao Tu[1], Zhihong Li [1] ✉ & Zhongming Zhao [2,5,6] ✉

Advances in mass spectrometry accelerates the characterization of HLA ligandome, necessitating the development of efficient methods for immuno-peptidomics analysis and (neo)antigen prediction. We develop ImmuneApp, an interpretable deep learning framework trained on extensive HLA ligand datasets, which improves the prediction of HLA-I epitopes, prioritizes neoe-pitopes, and enhances immunopeptidomics deconvolution. ImmuneApp extracts informative embeddings and identifies key residues for pHLA binding. We also present a more accurate model-based deconvolution approach and systematically analyzed 216 multi-allelic immunopeptidomics samples, iden-tifying 835,551 ligands restricted to over 100 HLA-I alleles. Our investigation reveals the effectiveness of the composite model, denoted as ImmuneApp-MA, which integrates mono- and multi-allelic data to enhance predictive perfor-mance. Leveraging ImmuneApp-MA as a pre-trained model, we built Immu-neApp-Neo, an immunogenicity predictor that outperforms existing methods for prioritizing immunogenic neoepitope. ImmuneApp demonstrates its utility across various immunopeptidomics datasets, which will promote the dis-covery of novel neoantigens and the development of new immunotherapies.

The adaptive immune system is capable of recognizing and killing the infected and malignant cells that present non-self and aberrant pep-tides by cytotoxic T cell receptors binding to antigens[1-3]. In humans, among all the factors needed for CD8 + T lymphocytes to initiate an immunogenic reaction, the critical gatekeeping step is the availability of peptides bound by human leukocyte antigen (HLA) class I mole-cules; this phenome has the potential for many clinical applications, such as malignant tumors, and cancer immunology[4-8]. For instance, tumor-specific antigens (neoantigens), which are newly produced in the cells due to events such as somatic mutations, alternative splicing, gene fusion, and viral infection, are pivotal in the process of immu-noediting and represent an important class of anticancer therapeutic targets for cytotoxic T cells[9-13]. While neoantigens can trigger a potent anti-tumor immune response, personalized immunotherapy devel-oped against them has now been used in clinical trials in a variety of solid tumors[6,14-16]. Moreover, understanding the specific epitopes of SARS-CoV-2 that are targeted by T cells through their T-cell receptors (TCRs) will facilitate peptide-based vaccine development[17-19]. There-fore, the identification and characterization of the landscape of the peptides exhibited by HLA-I molecules is an urgent need.

Early experiments conducted both in vitro and in vivo were designed to elucidate the binding characteristics of diverse peptide-HLA pairs, revealing allele-specific motifs within the presented peptides[20-25]. In vitro, studies have been primarily on quantifying the

[1]Department of Orthopaedics, The Second Xiangya Hospital, Central South University, Changsha, Hunan 410011, China. [2]Center for Precision Health, School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA. [3]Center for Advanced Parkinson Research, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [4]Genomics and Bioinformatics Hub, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA. [5]MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX 77030, USA. [6]Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA. [7]These authors contributed equally: Haodong Xu, Ruifeng Hu. ✉e-mail: haodong.xu@uth.tmc.edu; lizhihong@csu.edu.cn; zhongming.zhao@uth.tmc.edu

binding affinity between particular peptides and their corresponding HLA complexes through hypothesis-driven competitive binding assays[25]. Recent advancements in liquid chromatography and mass spectrometry (LC–MS/MS) techniques enhance the precision for extracting peptide-HLA (pHLA) complexes from cellular and tissue samples, enabling large-scale assays on the spectrum of peptides that are bound to HLA molecules[22,23,26–31]. Accordingly, the Human Immunopeptidome Project (HIPP) was formed to construct a comprehensive map of the human immunopeptidome. For example, Sarkizova et al. thoroughly profiled more than 185,000 peptides in across 95 mono-allelic cell lines for HLA-A, -B, -C, and -G. Their study substantially enhanced our understanding of HLA-associated peptidome in humans and illuminated the variety and complexity of endogenous HLA ligands[22]. Moreover, many clinical immunopeptidomics studies have been conducted recently for cancer immunotherapy[26,32–35]. The adoption of a multi-omics integration strategy, which combines immunopeptidomics, transcriptomics, and ribosome profiling, has emerged as a powerful approach for identifying numerous canonical or non-canonical tumor antigens tailored to individual patients[15,26,28,29,36–39]. This integrated approach aids in the advancement of customized anti-cancer or anti-virus vaccines with the potential to target patient-specific antigenic profiles, thus enhancing the efficacy of immunotherapeutic interventions.

In parallel with these technological advancements, computational approaches for pHLA binding prediction have undergone rapid growth. Initially, the methods have been developed only using MHC-peptide binding affinity data[40–43]. With the rapidly growing immunopeptidomics data, investigators have integrated such datasets into their prediction models[22,24,44–51]. Several initiatives leverage the unambiguous nature of mono-allelic ligands associated with precisely defined HLA alleles from genetically engineered cells[22,46]. Moreover, noteworthy contributions have been made by Gfeller et al., Bassani-Sternberg et al., Morten Nielsen et al., and Bulik-Sullivan et al., alongside other research entities, towards the utilization of multi-allelic immunopeptidomics datasets, i.e., comprising peptides corresponding to multiple cognate HLA alleles, to advance antigen presentation prediction[24,44,45,47,52,53]. Recent computational algorithms, including clustering-based deconvolution[45], iterative assignment[47], and direct modeling[24], have been released with the capacity to attribute multi-allelic ligands to individual MHC restrictions. For example, MixMHCp is capable of deconvoluting and assigning MHC restrictions through an unsupervised method[45].

Utilizing a mixture model algorithm, MixMHCp creates clusters of peptides and identifies binding motifs within a variety of datasets from MS with poly-specificity. Following this, it links each cluster to a specific HLA molecule, relying on the principles of co-occurrence and exclusion for HLA alleles. The NNAlign_MA algorithm uses mono-allelic ligands, which are restricted to specific HLA molecules, to pretrain a pan-specific prediction model[47]. This model then predicts all possible HLA molecules for a given ligand and determines HLA molecule restriction by identifying the highest rescaled prediction value. These methods have facilitated the expansion of training datasets and deepened our comprehension of the underlying principles governing pHLA binding and presentation. Through the integration of both multi-allelic and single-allelic ligands, predictors such as NetMHCpan4.1[49], MixMHCpred 2.2[50] and MHCflurry-2.0[51], and other tools have been developed and widely adopted. Additionally, certain algorithms are designed to exclusively model MHC-peptide binding, while others broaden their focus to incorporate more determinant factors of antigen processing and its presentation on the cell surface. Despite discrepancies regarding the optimal approach to modeling MHC-peptide binding, there exists a consensus that immunopeptidomics has enabled the generation of large-scale ligands, thereby substantially enhancing approaches for the prediction of peptides bound HLA molecules. Computational prediction of antigen presentation holds promise for identifying neoantigens and accelerating immunogenicity evaluation. However, most neoantigens lack immunogenic properties.[54,55]. Additionally, experimentally confirming immunogenicity requires considerable resources; it is crucial for (neo) antigen prediction methods to prioritize a significant proportion of immunogenic candidates among their top-ranked predictions. This prioritization is essential because only a limited number of top-ranked candidate neoantigens are subjected to clinical testing and practical application.

The growing immunopeptidomics datasets have formed an expanding repertoire of annotated HLA-associated peptides. Consequently, the development of efficient tools for deciphering immunopeptidomics data and more robust (neo)antigen presentation predictors is urgently needed. To tackle these challenges, we introduced ImmuneApp, a robust computational tool designed to facilitate the prediction of antigen presentation, assessment of neoepitope immunogenicity, and comprehensive immunopeptidomics analysis, all with heightened precision (Fig. 1). ImmuneApp leverages an interpretable, attention-based hybrid deep learning framework specifically designed for predicting HLA-I epitopes, trained on a dataset comprising 349,650 ligands. ImmuneApp enables the extraction of informative embeddings and the identification of critical residues governing peptide-HLA (pHLA) binding specificity. Thorough assessments conducted on independent mono-allelic datasets revealed that ImmuneApp outperforms current methods for predicting antigen presentation. Furthermore, we developed a more accurate, model-based deconvolution method and applied it to 216 publicly available multi-allelic immunopeptidomics samples, deconvoluting 835,551 ligands restricted to 104 distinct HLA-I alleles. We then evaluated the efficacy of a composite model, namely ImmuneApp-MA, which integrates both mono and multi-allelic data types to enhance predictive performances. To enhance our tool in clinical settings, we employed ImmuneApp-MA as a pre-trained model for deep transfer learning on a newly curated immunogenicity training dataset, resulting in the development of a novel immunogenicity predictor called ImmuneApp-Neo. Remarkably, ImmuneApp-Neo exhibited a substantially higher positive predictive value (PPV), with a 2.1-fold improvement when compared to all other models, in the identification of immunogenic neoepitopes. Finally, we developed an online platform (https://bioinfo.uth.edu/iapp/) to facilitate multitasking functionalities, including antigen presentation prediction, immunogenicity assessment, and analysis of immunopeptidomics cohort datasets. We applied ImmuneApp to the disease-related immunopeptidomics datasets sourced from tumor tissues and cancer biopsies. This application showcased the robustness of our method in various tasks such as quality control, binding annotations, HLA assignment, motif discovery and decomposition, and antigen presentation prediction in a sample-specific manner.

Results

Deep learning modeling pHLA presentation based on single-allelic eluted ligands

Using our collected single-allelic MS-eluted ligands data (Supplementary Data 1), we initially built the ImmuneApp-EL model to estimate the probability of a peptide's presentation by HLA-I molecules. This initial model was trained on 349,650 ligands covering 149 distinct HLA-I alleles, as well as 17,482,200 (50-fold excess) random peptides. Briefly, ImmuneApp-EL took encoded matrixes of the peptides and pseudo-sequences of HLA alleles on the BLOSUM50 substitution matrix. Then, the input matrixes were fed into a convolutional neural network (CNN) long short-term memory (LSTM), and attention modules for training. To accurately capture the inherent properties of peptide-HLA binding during training, features obtained from different parts of the neural network were retrieved from various layers, and then they were combined (Fig. 1a). To address the sample imbalance issue and enhance the
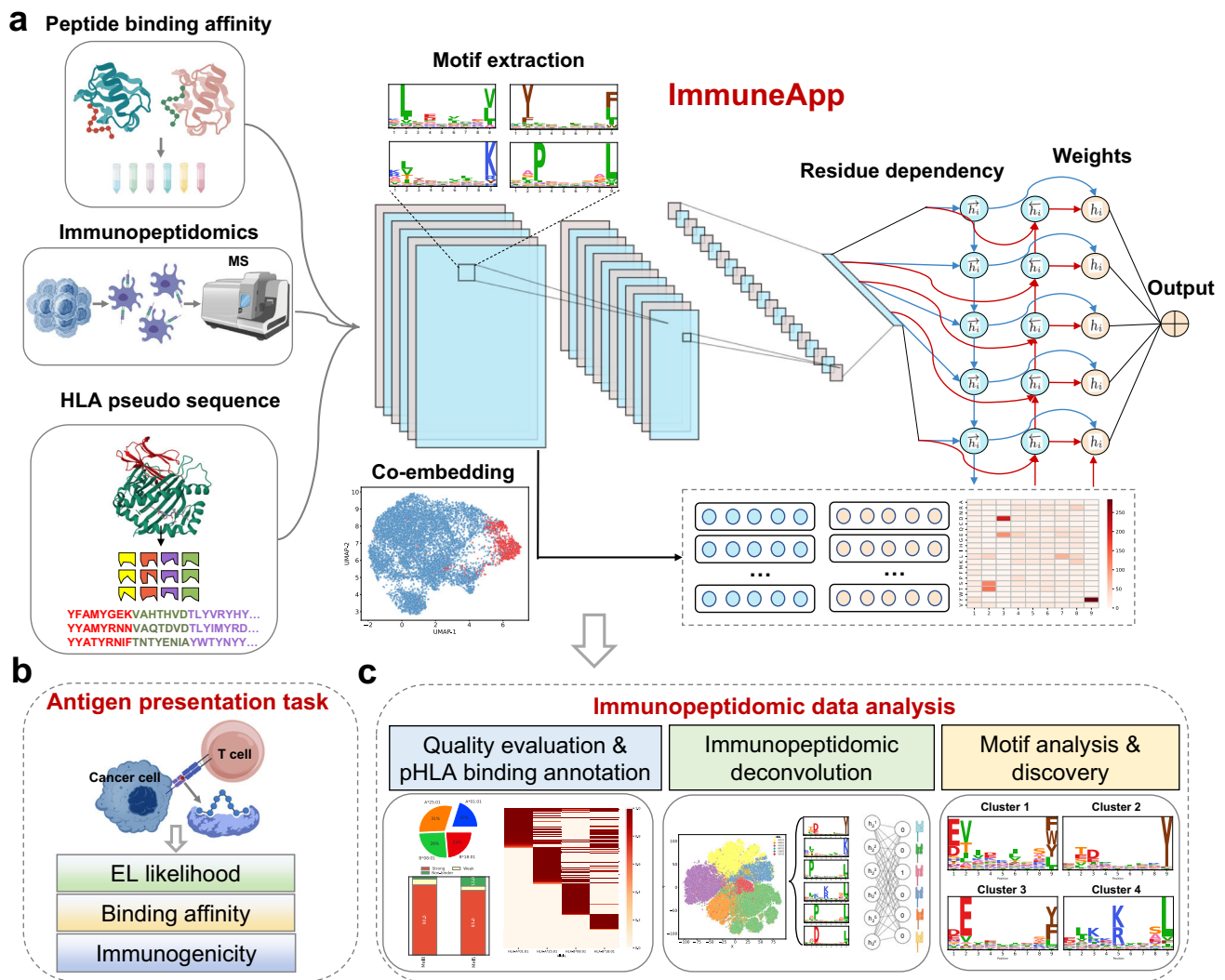
Fig. 1 | The framework, model architecture, and utility of ImmuneApp for HLA-I antigen prediction and immunopeptidome analysis. a The deep learning framework for prediction of human leukocyte antigen (HLA) class I antigen presentation. ImmuneApp took encoded matrixes of the peptides sourced from mass spectrometry-eluted HLA ligands or peptides with binding affinity (BA) measurements, and pseudo-sequences of HLA alleles on the BLOSUM50 substitution matrix. Then, the input matrixes were fed into a convolutional neural network (CNN) and long short-term memory (LSTM) with attention modules for training. Features obtained from different parts of the neural network were retrieved from various layers, and then they were combined. A probability of the likelihood of antigen presentation in the setting of certain HLA class I alleles is produced by the output layer that implements a sigmoid nonlinear transformation. **b** ImmuneApp provides various presentation prediction capabilities, including eluted ligand (EL) likelihood estimate, in vitro BA measurements, and immunogenicity prediction. **c** ImmuneApp provides one-stop analysis, statistical reports, and visualization for immunopeptidomics data, such as quality control, binding annotations, HLA assignment, motif discovery and decomposition, and antigen presentation prediction on a sample-specific basis.

robustness of the model, ImmuneApp-EL was implemented using a balanced class-weight approach and the ensemble learning strategy, resulting in the creation of 25 different models (see "Methods"). These models output an average score, representing the likelihood of the query ligand binding to the HLA molecule.

To evaluate the accuracy and robustness of our method, we compared ImmuneApp-EL with six peer methods using an external single-allelic set of 43,866 ligands (Supplementary Data 2). These methods are NetMHCpan-4.1 (rank and score), MixMHCpred-2.1 and 2.2 (rank and score), MHCflurry-2.0 (rank and score), HLAthena (rank and score), TransPHLA, and MHCnuggets-2.4. To ensure the orthogonality of the independent test set, we excluded any ligands that overlapped with the training data from the test dataset. A total of 2,471,337 random peptides served as negative data. We calculated three evaluation metrics: the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), and the PPV. Our findings, illustrated in Fig. 2a and Fig. S1,

demonstrated that ImmuneApp-EL enhances EL predictive performance, achieving a mean AUROC of 0.9576 and a mean AUPRC of 0.6139 when stratified by HLAs. The top-performing method among previous approaches was MHCflurry-2.0, which recorded a mean AUROC of 0.9370 (score) and a mean AUPRC of 0.5334 (rank). Additionally, the median PPV values across alleles for each method were as follows: 0.7656 for ImmuneApp-EL scores, 0.7085 for MHCflurry-2.0 (both ranks and scores), 0.6883 for NetMHCpan-4.1, 0.6703 for HLAthena ranks, 0.6538 for MixMHCpred-2.2 scores, 0.5806 for MixMHCpred-2.1 scores, 0.6077 for TransPHLA, and 0.4631 for MHCnuggets-2.4. ImmuneApp-EL excelled in distinguishing MS hits from decoy peptides based on PPV values. Furthermore, we enhanced our evaluation by implementing a more detailed stratification that considers both HLA and peptide length. In this analysis, ImmuneApp-EL once again surpassed all other tools, achieving mean AUROC and AUPRC values of 0.9406 and 0.5820, respectively. In contrast, the top-performing method among its peers, MHCflurry-2.0, recorded average
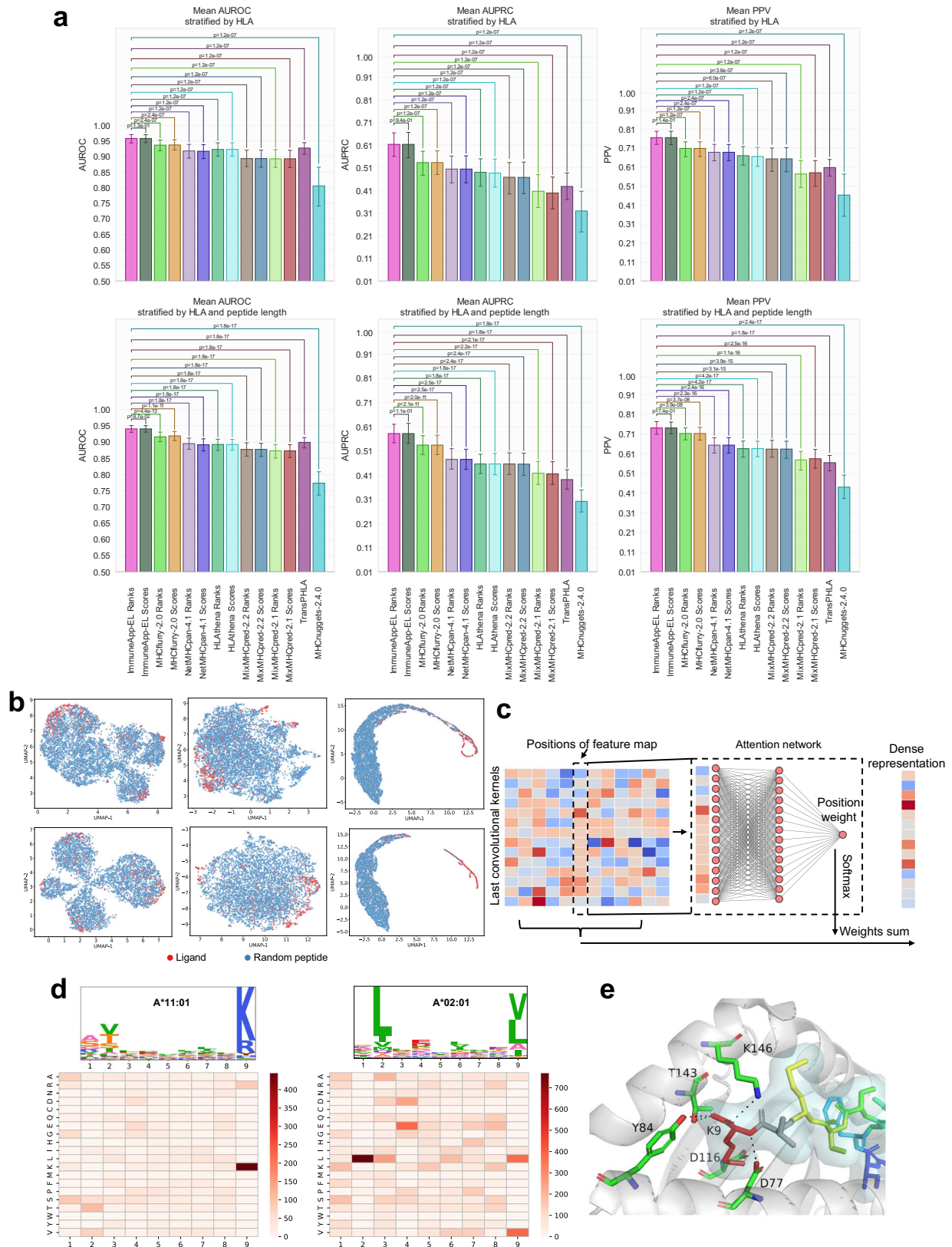
Fig. 2 | The validation performance and model interpretability of ImmuneApp.
a Benchmark comparison of ImmuneApp, NetMHCpan-4.1, MHCflurry 2.0, MixMHCpred 2.1&2.2, HLAthena, MHCnuggets-2.4, and TransPHLA. Mean AUROC, AUPRC, and PPV with 95% confidence interval (CI) stratified by HLA ($n = 24$) and both HLA and epitope length ($n = 96$) were calculated, followed by two-tailed Wilcoxon signed-rank test to estimate adjusted P-values. **b** Uniform Manifold Approximation and Projection (UMAP) visualization of ligands (red) and random peptide (blue) representations in different layers for HLA alleles of A*02:01 (upper) and A*11:01 (lower). **c** The attention architecture implemented in the ImmuneApp. **d** Sequence motifs and accumulated attention scores for the peptides binding to HLA alleles of A*11:01 (left) and A*02:01 (right). **e** Structural data (PDB ID: 3RL1) supports the motif revealed by ImmuneApp for peptide presented by HLA-A*11:01. Source data are provided as a Source Data file.

AUROC and AUPRC values of 0.9189 and 0.5343, respectively. ImmuneApp-EL also excelled compared to all other methods in terms of PPV. Overall, these results confirm that our model outperforms prior tools, demonstrating its efficacy in recognizing presented antigens.

ImmuneApp facilitates informative embeddings and captures pHLA binding motif

The hidden layers in ImmuneApp enable the mapping of peptides to a lower dimensional representation space, which could be extracted and visualized. Thus, we visualized the numeric embeddings of ligands and random peptides for several well-characterized HLA alleles (Fig. 2b, Fig. S2), such as A*02:01 and A*11:01. Our model demonstrated the ability to hierarchically capture a more effective and interpretable representation of pHLA complexes. Specifically, at the input layer, the extracted features for ligands and random peptides were mixed. However, as the predicted features passed through the deep learning framework, the model began differentiating between ligands and random peptides. In the fully connected layer before the output, we found the ligands and random peptides could be separated, and they were grouped into two distinct clusters by the low-dimensional projection. The results proved that the deep learning-based embeddings could well represent the information on binding specificities of pHLA complexes.

Furthermore, we elucidated black boxes of deep learning by interpreting the weights learned by the attention layer in the ImmuneApp (Fig. 2c). Importantly, consistent amino-acid preferences were observed at critical peptide positions for several HLA alleles (Fig. 2d, Fig. S3) whose binding motifs have been illustrated in Motif Viewer of NetMHCpan[56] or the MHC Motif Atlas[57]. For example, the amino acid lysine (Lys, K) at position 9 was characterized as an anchor residue with the highest weight for the peptides binding to HLA-A*11:01. Structural analyses revealed that the lysine residue at the C terminus participated in a dense network of hydrogen bonds with neighboring residues in the bound HLA molecule (PDB ID: 6JOZ). This finding underscores the biological significance of the binding motifs identified by ImmuneApp (Fig. 2e, Fig. S4). In summary, through interpretable, learned features and weights, we demonstrated that our model could automatically learn informative embeddings and capture critical amino acids that contributed to defining HLA-binding motifs during the training.

Model-based approach improves multi-allelic immunopeptidomics deconvolution

Immunopeptidome data present significant complexity due to the presence of multiple HLA alleles on the cell membrane. A major challenge in modeling and analyzing multi-allelic data is the deconvolution of immunopeptidomics, which involves assigning a ligand to its corresponding allele. By decoding the weights learned by the attention layer, we have shown that the mono-allelic model (ImmuneApp-EL) exhibits high accuracy and robustness, enabling effective capture of binding motifs. Considering the superior performance of the trained model, here we here introduced a model-driven deconvolution method to transform immunopeptidomics data into pseudo-mono-allelic ligands. Briefly, we made predictions for all HLA alleles annotated for each sample, and raw scores were calibrated using percent rank values against a reference set of 500,000 random peptides. Subsequently, for each sample, we excluded allele-peptide pairs with a predicted binding rank greater than 20% to remove potential contaminants. Additionally, we selected the allele-peptide pair with the lowest rank, signifying the strongest binding affinity (the best binder), while disregarding all other pairs (Fig. 3a).

Our method was compared with NetMHCpan4.1 and MixMHCpred 2.2, which utilize NNalign-MA and MixMHCp for the deconvolution of immunopeptidomics data, respectively. We curated a dataset of 435,397 eluted ligands covering 86 HLA alleles from 47 recently published samples (Supplementary Data 3). Initially, AUROC, AUPRC, and PPV were computed to evaluate the capacity of predictors in recognizing true ligands within extensive random peptide libraries. Our method, as illustrated in Fig. 3b and Fig. S5, enhanced EL predictive performance, achieving a mean AUROC of 0.9650 and a mean AUPRC of 0.7600 when stratified by samples. In contrast, NetMHCpan-4.1 yielded a mean AUROC of 0.9155 and a mean AUPRC of 0.6071, while MixMHCpred-2.2 produced a mean AUROC of 0.9029 and a mean AUPRC of 0.6328. Notably, the PPV values across samples for each method were 0.8747 for our approach, 0.7689 for NetMHCpan-4.1, and 0.7970 for MixMHCpred-2.2. These results indicated that our method is more effective in identifying HLA-bound peptides present in patient-derived tumor tissues or cell lines. We further enhanced our evaluation by implementing a more detailed stratification, considering both sample and peptide length. Once again, our method surpassed the other two tools, achieving mean values of 0.9239 for AUROC, 0.6410 for AUPRC, and 0.7913 for PPV (see Fig. 3c). In contrast, NetMHCpan-4.1 recorded a mean AUROC of 0.8550, a mean AUPRC of 0.5080, and a mean PPV of 0.6750. Similarly, MixMHCpred 2.2 obtained a mean AUROC of 0.8518, a mean AUPRC of 0.5367, and a mean PPV of 0.7103. Compared to these two well-established tools trained on immunopeptidomics data, our approach demonstrated improvements of 8.06%, 19.43%, and 11.40% in AUROC, AUPRC, and PPV, respectively. We further explored the congruence between HLA binding motifs derived from deconvolution and motifs identified by single-allelic ligands. The average Pearson's correlation coefficient (PCC) was calculated across alleles (Fig. 3d). The motifs for HLA-I alleles identified in mono-allelic data showed a high degree of similarity to those found in multi-allelic samples, and our method achieved higher average PCC values among alleles when compared to NetMHCpan4.1 and MixMHCpred 2.2, indicating its capability for accurately assigning a peptide to its cognate allele in multi-allelic data. The improvement is statistically significant (adjusted $P < 0.05$, two-tailed Wilcoxon signed-rank tests) (Fig. 3c, d).

Integrating large-scale immunopeptidomes improves the prediction of antigen presentation

Considering the abundance of publicly accessible immunopeptidomics data from various tumor samples and cell lines, we posited that systematically integrating all high-quality datasets would enhance the pHLA representation in the training data and lead to more accurate predictions. To achieve this, we further collected and processed 216 publicly available multi-allelic immunopeptidomics samples. The final curated dataset comprises 969,435 ligands restricted to 110 HLA-I molecules (Supplementary Data 4). Employing our model-driven deconvolution method, we transformed immunopeptidomics data into pseudo-single-allelic ligands, successfully mapping 835,551 ligands (86.19%) to 104 alleles. The motifs for HLA-I alleles identified in mono-allelic data showed significant similarity to those in multi-allelic samples (Fig. S6), indicating that our approach effectively assigns peptides to their corresponding alleles in multi-allelic data. After removing duplicates, we obtained a total of 328,227 unique allele-ligand pairs. Subsequently, we merged the mono-allelic and pseudo-mono-allelic data, resulting in 573,453 unique allele-ligand pairs covering 162 alleles (Supplementary Data 5). This final dataset was utilized to train a mixed prediction model, ImmuneApp-MA.

We benchmarked the performance of ImmuneApp-MA against all other tools, including ImmuneApp-EL, while allele-ligand pairs overlapping with the new training data were removed from testing data. As illustrated in Fig. 4a and Fig. S7, ImmuneApp-MA achieved mean values of 0.9496 for AUROC, 0.4710 for AUPRC, and 0.6347 for PPV. It outperformed ImmuneApp-EL in AUROC, AUPRC, and PPV, and significantly surpassed other methods, including NetMHCpan-4.1, MixMHCpred-2.1 and 2.2, MHCflurry-2.0, HLAthena, TransPHLA, and MHCnuggets-2.4. Among these peer methods, MHCflurry-2.0
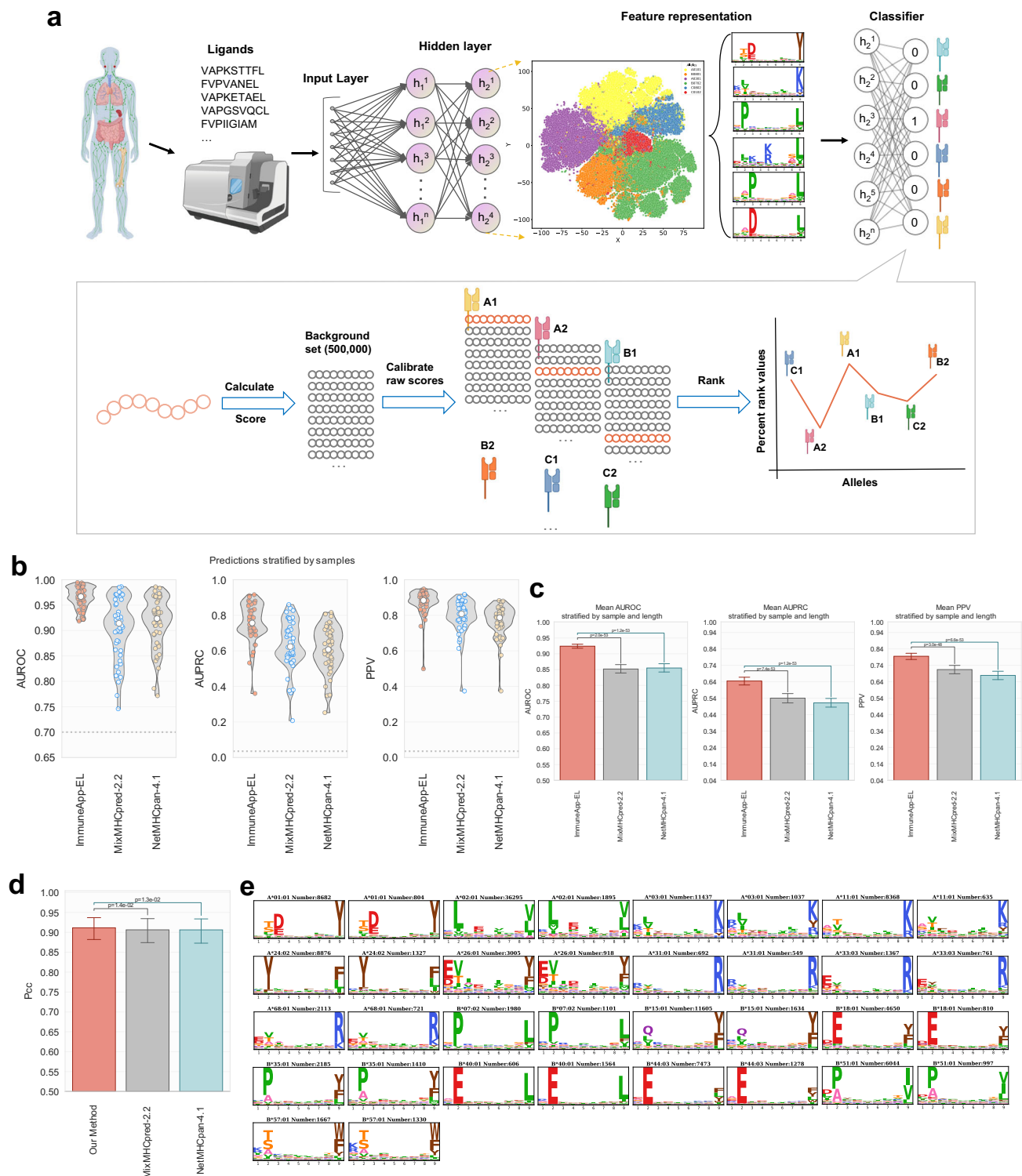
Fig. 3 | The framework and performance of model-based approach for multi-allelic immunopeptidomics deconvolution. a The framework of the deconvolution method to convert multi-allelic immunopeptidomics data into pseudo-mono-allelic. Using a mono-allelic model, predictions were made for all HLA alleles (up to six) of each sample. To make the predicted scores for different alleles comparable, raw scores were calibrated using percent rank values from a background set of 500,000 random peptides. Peptide was assigned to the allele with the lowest rank (best binder). **b** Benchmark comparison of our approach, NetMHCpan-4.1 and MixMHCpred 2.2, to retrieve HLA-bound peptides observed in patient-derived tumor datasets. AUROC, AUPRC, and PPV stratified by samples ($n = 47$) were calculated. **c** Mean AUROC, AUPRC, and PPV values with 95% confidence interval (CI) stratified by both samples and epitope length ($n = 317$) were calculated, followed by a two-tailed Wilcoxon signed-rank test for adjusted P-values. **d** Average Pearson's correlation coefficient (PCC) among alleles ($n = 24$) for HLA binding motifs identified by single-allelic ligands and revealed by our approach, NetMHCpan-4.1, and MixMHCpred 2.2. Two-tailed Wilcoxon signed-rank test was used for the calculation of P-values. Bars represent means and error bars are 95% CIs. **e** Sequence logos of binding motif for HLA-1 alleles revealed by our approach from 47 multi-allelic immunopeptidomics samples compared to that from external mono-allelic ligands. Among the two logos of the same HLA allele, the left one was obtained by deconvolution, and the right one originated from mono-allelic ligands. Source data are provided as a Source Data file.
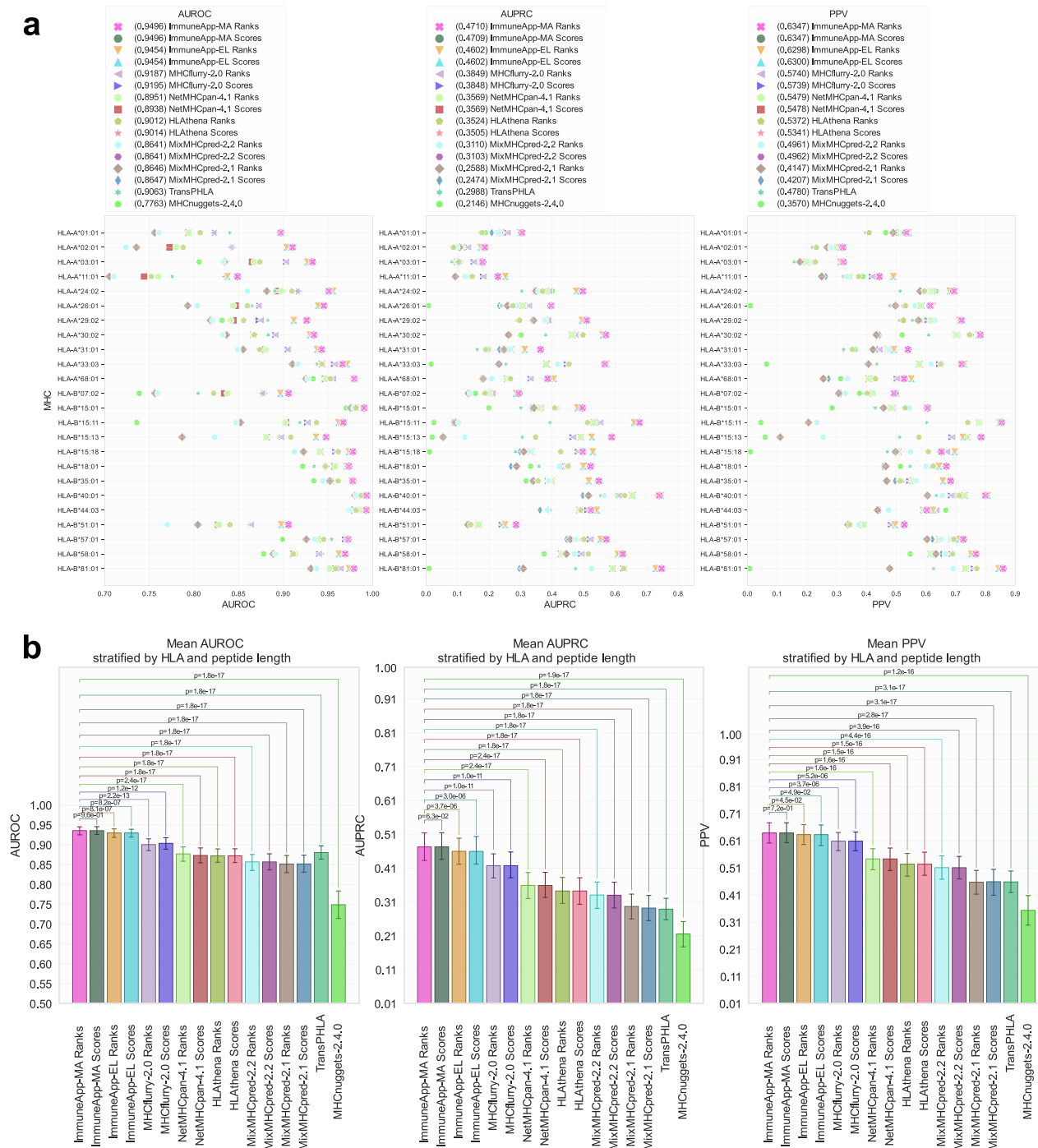
Fig. 4 | Benchmark comparison of composite model ImmuneApp-MA that integrates both mono and multi-allelic immunopeptidomics data. a. AUROC, AUPRC, and PPV values for each allele in the external mono-allelic EL testing dataset for ImmuneApp-MA, ImmuneApp-EL, NetMHCpan-4.1, MHCflurry 2.0, MixMHCpred 2.1&2.2, HLAthena, MHCnuggets-2.4 and TransPHLA. **b**. Mean AUROC, AUPRC, and PPV values with a 95% confidence interval (CI) stratified by both HLA and epitope length ($n = 96$) were calculated, followed by a two-tailed Wilcoxon signed-rank test for adjusted P-values. Source data are provided as a Source Data file.

performed best, with a mean AUROC of 0.9195 (score), a mean AUPRC of 0.3849 (rank), and a mean PPV of 0.5740 (rank). We further stratified the results by HLA alleles and peptide length, as shown in Fig. 4b. ImmuneApp-MA yielded mean AUROC and AUPRC values of 0.9355 and 0.4704, respectively, while the top peer method, MHCflurry-2.0, produced a mean AUROC of 0.9036 (score) and a mean AUPRC of 0.4156 (rank and score). ImmuneApp-MA also demonstrated superior PPV compared to all other peer predictors. Overall, ImmuneApp-MA surpassed all peer methods with this more detailed stratification, and

the improvement was statistically significant (adjusted $P < 0.05$, two-tailed Wilcoxon signed-rank test, see Fig. 4b and Fig. S7).

Transfer learning strategy enhances neoepitopes immunogenicity screening
One significant application of antigen presentability lies in the identification of neoepitopes, which offers considerable potential for immunotherapies. Neoepitope-directed therapy represents a promising strategy aimed at harnessing the host immune response against

tumor neoantigens to eliminate cancer cells. Neoepitopes-directed therapy is a promising strategy for targeting the host immune response against tumor neoantigens to eradicate cancer cells. Accurate predictions of immunogenic neoepitopes are essential for identifying potential targets for immunotherapeutic strategies, including adoptive cell therapy and peptide-based vaccines. Nonetheless, most neoantigens lack immunogenic properties, and there is currently no high-throughput method available for experimentally verifying their immunogenicity. Therefore, it is essential for (neo)antigen prediction methods to emphasize a considerable proportion of immunogenicity within their highest-ranking prediction scores, as only a select few candidate neoantigens ranked at the top undergo clinical testing and practical application. Using a carefully curated neoepitope dataset consisting of 349 immunogenic and 1838 non-immunogenic neoepitopes (Supplementary Data 6), we first assessed the capability of ImmuneApp-MA for prioritizing immunogenic neoepitopes and compared it to nine peer tools. The PPV values calculated for these methods were ImmuneApp-MA (0.3720), ImmuneApp-EL (0.3595), MHCflurry 2.0 (0.3356), NetMHCpan-4.1 (0.3313), PRIME 2.0 (0.3200), PRIME 1.0 (0.2418), MixMHCpred-2.2 (0.3222), HLAthena (0.3063), TransPHLA (0.2654), and MHCnuggets-2.4.0 (0.2154). ImmuneApp-MA had the best performance and improved the performance of immunogenic neoepitope prediction.

The current availability of neoepitopes capable of eliciting a clinically detectable antigen-specific immune response is restricted. Motivated by the growing utility of big pre-trained models in biomedical multi-omics analysis in recent years, we investigated the potential enhancement of neoantigen immunogenicity prediction through the utilization of deep transfer learning. We applied the ImmuneApp-MA as pre-trained model to employ deep transfer-learning on a new curated immunogenicity training data, resulting in the creation of a novel immunogenicity predictor named ImmuneApp-Neo (Fig. 5a). To evaluate the prediction performance of ImmuneApp-Neo, we calculated the PPV among the top n outputs (PPVn), which indicates the proportion of the top n pHLAs that are genuinely immunogenic. Additionally, we computed other metrics such as AUROC and AUPRC. By plotting PPVn against all numbers of neoepitopes, our analysis demonstrated that ImmuneApp-Neo outperformed all other predictors in immunogenic neoepitope prediction (Fig. 5b). To summarize this PPVn curve, the mean PPVn was illustrated with 95% confidence interval (CI) whiskers in Fig. 5c, revealing that ImmuneApp-Neo attained a mean PPV of 0.7151, significantly surpassing the best peer method, MHCflurry 2.0, which only achieved a mean PPV of 0.3356. This result underscores the effectiveness of transfer learning for immunogenicity prediction, as ImmuneApp-Neo notably outperformed ImmuneApp-MA, which had a mean PPV of 0.3711. Furthermore, ImmuneApp-Neo demonstrated higher AUROC and AUPRC values of 0.7928 (Fig. 5d) and 0.5542 (Fig. 5e), respectively, outperforming all peer methods. Collectively, ImmuneApp-Neo exhibited a substantially greater PPV, with a 2.1-fold improvement compared to other models in identifying immunogenic neoepitopes. Our study highlights the importance of integrating and utilizing large-scale immunopeptidome data from clinical samples to investigate immunogenicity in a more relevant physiological context.

Discussion

In recent years, extensive MS-eluted MHC ligand data have become available, allowing for a comprehensive characterization of the MHC-presented ligandome. Leveraging these datasets, we created ImmuneApp, a versatile toolbox designed for predicting and analyzing antigen sequences presented on specific HLA-I types. ImmuneApp can be employed for three primary tasks (Fig. 1b, c). First, by integrating large-scale immunopeptidomics data, we developed a novel pan-specific algorithm named ImmuneApp-MA to enhance the accuracy of HLA-I antigen presentation predictions. Second, by decoding the

weights learned by the attention layer, we show the deep-learning framework enables accurate capture of binding motifs and thus developed a model-based method to improve the deconvolution analysis of immunopeptidomics for assigning multi-allelic ligandome to cognate alleles. Third, by employing ImmuneApp-MA as a pre-trained model for deep transfer learning on the curated immunogenicity dataset, we created a novel immunogenicity predictor named ImmuneApp-Neo, which outperformed nine leading methods in the identification of immunogenic neoepitopes. Collectively, by broadening the training datasets and refining the algorithms, our ImmuneApp demonstrated enhanced prediction capabilities for both HLA-I ligands and neoepitopes.

The quality and depth of training data are two crucial factors in computational models. Benefiting from our collected 349,650 mono-allelic ligands, we developed an accurate model-driven deconvolution method for reprocessing 216 immunopeptidomics samples, successfully deconvoluting 835,551 ligands. The pseudo-mono-allelic ligands were integrated, resulting in an expansion of the training set size. In our composite model (ImmuneApp-MA), integration of both data types could enhance predictive performance. Specifically, ImmuneApp-MA demonstrated the highest predictive capability, notably surpassing seven peer methods across various HLA loci and peptide lengths. In addition, we explored the interpretability of the model. Most neural networks for predicting molecular properties and biological activities act as black boxes. However, ImmuneApp is easily interpretable because it implements an attention-based hybrid deep-learning framework for predicting HLA-I epitopes. The hidden layers in ImmuneApp enabled the mapping of peptides to a lower-dimensional representation space that could be extracted and visualized. Our results indicated that the deep learning-based embeddings effectively represented the binding specificities of pHLA complexes. Furthermore, by decoding the learned features and weights, we demonstrated that our framework could effectively extract interpretable patterns. This feature is supported by existing structural data, and our analysis validated the biological significance of the binding motifs identified by ImmuneApp. Of note, this analysis does not aim to replicate already established and well-characterized HLA-I binding motifs. While some prediction algorithms, such as MixMHCpred and those based on position weight matrices (PWMs), are designed for easy interpretability in motif discovery, our method differs by offering additional layers of interpretability beyond traditional PWM-based methods. Our model could automatically learn and identify critical amino acids that define HLA-binding motifs during training, providing both interpretability and predictive capabilities essential for immunopeptidomics analysis and HLA-I ligand prediction.

An accurate prediction of immunogenic neoepitopes helps identify potential targets for the design of immunotherapeutic strategies, such as adoptive cell therapy and peptide-based vaccines. While the primary objective of (neo)antigen prediction is to screen for valid neoepitopes that elicit a clinically relevant antigen-specific immune response, the availability of immunogenicity data for training deep learning models is limited. To tackle this data scarcity issue and enhance the effectiveness of our tool in clinical applications, we initially trained the base models on presentation data and then employed transfer learning using immunogenicity data to create ImmuneApp-Neo. We subsequently evaluated ImmuneApp-Neo alongside nine related methods using an independent dataset focused on neoepitope immunogenicity. Our evaluation revealed that ImmuneApp-Neo demonstrated high precision, significantly surpassing other methods in immunogenicity prediction. Moreover, we observed that even when transfer learning techniques were not used, ImmuneApp-MA was still superior to other methods, suggesting that integration and utilization of large-scale immunopeptidomes data from (clinical) samples is essentially needed. This infers the use of the model to study immunogenicity in a more relevant and physiological context.
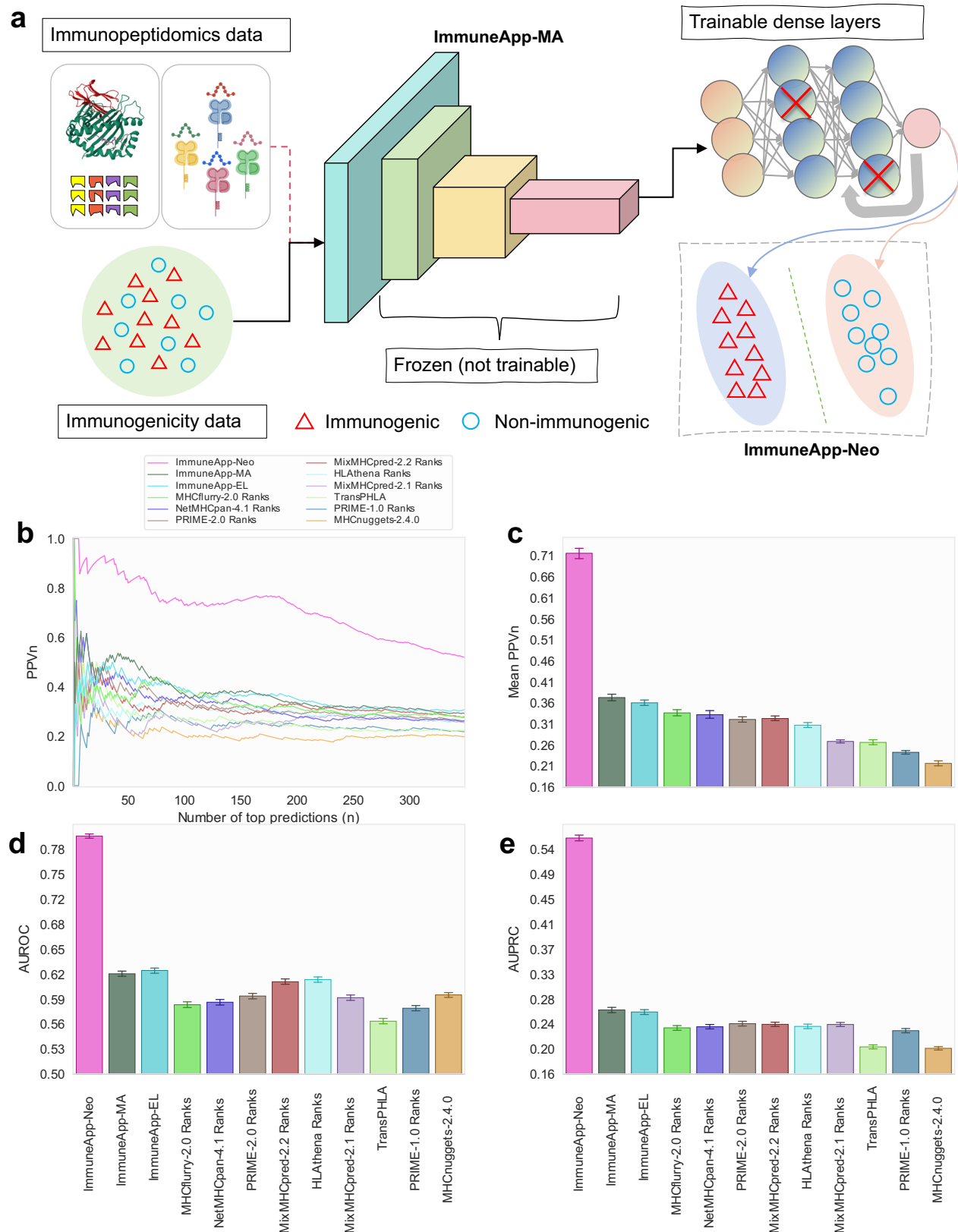
Fig. 5 | The transfer-learning framework and validation performance of ImmuneApp-Neo for immunogenicity prediction. a The development of ImmuneApp-Neo with a transfer-learning strategy for immunogenicity prediction. ImmuneApp-Neo was trained on new curated immunogenicity data by retraining the last three fully connected layers of the mixed prediction model ImmuneApp-MA, which outputs the neoepitope immunogenicity score. **b** PPVn was calculated for all benchmark methods, including ImmuneApp-Neo, ImmuneApp-MA, ImmuneApp-EL, PRIME 1.0&2.0, NetMHCpan-4.1, MHCflurry 2.0, MixMHCpred 2.1&2.2, HLAthena, MHCnuggets-2.4, and TransPHLA, as the fraction of neoepitopes that are immunogenic within the top n predictions (value of n ranges from 1 to 349). **c** Mean PPVn with a 95% confidence interval (CI) for all methods are shown. It summarizes the PPVn curves for all valid choices of n ($n = 349$). **d, e** Mean AUROC (**d**) and AUPRC (**e**) were calculated for all benchmark methods for neoepitope immunogenicity prediction. Bars represent means and error bars are 95% CIs. Source data are provided as a Source Data file.

To make ImmuneApp easy to use, we developed an online platform (https://bioinfo.uth.edu/iapp/) with multi-tasking functionalities, including antigen presentation prediction, immunogenicity assessment, and immunopeptidomics data analysis. We also implemented a module to predict the binding affinities between peptides and pan HLA-I molecules, trained by over 200,000 quantitative BA measurements (Fig. S8). We applied ImmuneApp to analyze multiple immunopeptidomics datasets from melanoma tumor tissues[14], lung, and gastric cancer biopsies[15] as case demonstration (Figs. S9–13), such as quality control, motif analysis, and pHLA binding and deconvolution. For example, ImmuneApp provided quality control analysis to determine the general quality of immunopepitdomic data. Our results indicated low-quality metrics of immunopeptidomic data in lung and gastric cancer biopsies, whereas three melanoma tumor tissues yielded high-quality immunopeptidomic data. Recent advances in pMHC profiling techniques have indeed reduced the minimum sample input requirements from an initial count of over 10^9 cells to 10^7 cells. However, this reduction does not completely eliminate the substantial limitation it imposes on the clinical applicability of the method. The paucity of material in clinical samples, such as those obtained through fine needle biopsies, often proves inadequate for extensive pMHC profiling. This demands the development of more sensitive and high-throughput technologies for the efficient detection of MHC-presented ligands. Moreover, we implemented both unsupervised gibbscluster and supervised allele-specific approaches for motif analysis. The former is a standard GibbsCluster[58] run using all peptides, while the latter is dependent on the results of our model-based deconvolution to assign multi-allelic data to cognate alleles. For large-scale immunopeptidomic data, installation of ImmuneApp on the local computer is recommended.

Our evaluation of ImmuneApp has provided promising results. It offers an alternative tool for HLA-I epitope prediction and immunopeptidome analysis. However, there are several limitations. First, ImmuneApp is restricted to operating solely on HLA-I data, while some other methods are capable of predicting both HLA-I and HLA-II presentation. We intend to augment our dataset by incorporating additional publicly available multi-allelic HLA-II immunopeptidomic data, and subsequently expand the tool's functionality for predicting and analyzing HLA-II antigen presentation. Second, in our assessment of immunogenicity, we only considered the probability of antigen presentation by HLA-I molecules. Although this parameter represents a pivotal determinant of immunogenicity, it is imperative to incorporate various other intrinsic factors to enhance the accuracy of immunogenicity prediction. Such factors include but are not limited to, proteasomal cleavage preferences, TAP transport efficiency, physicochemical properties such as hydrophobicity, and TCR recognition. Moreover, validation of predicted outcomes through additional experiments, encompassing both mass spectrometry assays and immunogenicity assays, is typically needed. Third, emerging evidence suggests a correlation between higher mutational burden and enhanced antitumor activity following CTLA4 or PD-1 blockade. The quantity of neo-peptides originating from somatic mutations, gene fusions, alternative splicing, and related factors, appears crucial for predicting the efficacy of immunotherapy[59]. Consequently, the development of robust methods for neoantigen selection under the selective pressure exerted by immune checkpoint blockade will likely be important for refining our approach for clinical applicability. Finally, inherent biases within MS data, such as the overrepresentation of "flyable" peptides, pose challenges to detectable ligand repertoire and consequently influence the acquired binding motifs. Consequently, to gain a more comprehensive understanding of HLA antigen presentation, advanced equipment specifically designed for high-throughput detection of MHC-peptide interactions is essential.

Methods

Mass spectrometry-eluted HLA ligands

Mono-allelic data. Single-allelic EL datasets were collected and processed from the training data of NetMHCpan-4.1[49] and MHCflurry-2.0[51], which was carefully processed and filtered from publications by Sarkizova et al.[22] and Abelin et al.[46] as well as MS hits from the IEDB[43], SysteMHC Atlas[60], and established datasets from their previous versions. Moreover, multiple HLA-I peptidomics from studies by Jappe et al.[61] and Faridi et al.[62] were obtained. These datasets were then integrated and duplicate entries were removed. All peptides employed in the new training dataset were filtered to only include 8 to 15 amino acid long peptides, resulting in 349,650 ligands restricted to 149 distinct HLA-I alleles. We referred these to the MONOALLELIC training data (Supplementary Data 1). To benchmark the predictors in this study, we collected an external single-allelic dataset from a recently published HLA-I peptidomics[63]. This dataset contained 43,866 HLA-I ligands; it was not included in the training of any previous predictors. This dataset was referred to as MONOALLELIC-testing data (Supplementary Data 2).

Multi-allelic data. To increase the number of ligands and encompass a wider array of HLA-I alleles, we incorporated publicly available multi-allelic HLA peptidomics data, where the precise HLA class I restrictions had not been experimentally established. Only samples with confirmed HLA-I typing were utilized. We categorized all curated samples into two groups. MULTIALLELIC-Recent included 47 samples from four recently published HLA-I peptidomics studies[35,63–65]. It contained 435,397 eluted ligands covering 86 different HLA alleles (Supplementary Data 3). This dataset was not used to train any previous predictors. Therefore, it was employed to benchmark the predictors developed in this study and others in a multi-allelic setting. This benchmarking involved assessing the predictors' performance in identifying true ligands within extensive random peptide libraries and comparing the consistency of HLA binding motifs derived through deconvolution with established motifs. MULTIALLELIC-All comprised 948,160 identified ligands from 216 samples representing 110 different HLA alleles (Supplementary Data 4, MUTLIALLELIC-Recent was included). This dataset was transformed into pseudo-mono-allelic data using our developed deconvolution method, which was then combined with the actual mono-allelic data to train a comprehensive model. Notably, all data were obtained from the original publications without being filtered by any HLA-I ligand predictors. This approach ensures that our dataset remains free from biases introduced by such filtering.

Quantitative binding affinity measurements

The most widely used dataset of MHC-I binding affinity was originally acquired from the IEDB[43]. To develop a model capable of predicting peptide binding to various MHC molecules, especially in humans, we also incorporated another dataset from Pearson et al.[66]. The final dataset included over 200,000 quantitative BA measurements across peptides and 190 MHC-I alleles (Supplementary Data 7). The IPD-IMGT/HLA database was used to retrieve the MHC molecule sequences. The following equation was used to convert the peptide–MHC binding affinities represented as IC50 in nM units:

$$\text{Transformed score} = 1 - \frac{\log(\text{Affinity measurement})}{\log(50000)} \qquad (1)$$

The neoepitope immunogenicity dataset

We extracted immunogenicity data from the training datasets of PRIME-1.0[55] and PRIME-2.0[50], as well as data obtained from the IEDB on December 19, 2023. The training datasets used from PRIME-1.0 and PRIME-2.0 included neoepitopes, viral antigens, and cancer-testis antigens. The first training datasets used in PRIME-1.0 and PRIME-2.0 included neoepitopes, viral antigens, cancer-testis antigens, and 9-mer

peptides that were chosen at random from the human proteome to serve as negative examples. Objectives of our queries to the IEDB included human hosts, infectious illnesses, T-cell tests, linear peptides, and MHC-I restriction of our queries to the IEDB included human hosts, infectious illnesses, T-cell tests, linear peptides, and MHC-I restriction. After eliminating overlapping data with all previously curated datasets, the final immunogenicity data included a total of 5050 immunogenic neoepitopes and 7745 non-immunogenic ones. We refer to these as the IMMUNOGENIC training data. We employed deep transfer learning on this immunogenicity data, resulting in the creation of a novel immunogenicity predictor. Further, to benchmark the immunogenicity predictors developed here, and others, an external immunogenicity dataset was compiled by collecting NEPdb[67], Neopepsee[68], TESLA[54], and the data from 16 cancer patients using the MANAFEST assay[69,70]. After excluding overlaps with all other single- and multi-allelic peptidomics, as well as the IMMUNOGENIC-training data, we compiled 349 immunogenic neoepitopes and 1838 non-immunogenic ones (see Supplementary Data 6, IMMUNOGENIC-testing data).

Decoy selection

The negative datasets were constructed by randomly picking peptides (decoys) in the UniProt human reference proteome (UP000005640_9606) that did not overlap with the identified ligands (hits). We constructed a pool of random peptides (8–15 amino acids long) and sampled a large number of length-matched decoy peptides with the observed allele to avoid bias. We excluded all peptides, including both hits and decoys, that contained non-canonical amino acids. Decoy generation for benchmarking purposes was conducted separately from the decoy generation employed during model training.

Peptide representation

MHC sequences and peptides are used as input by the ImmuneApp architecture. These sequences were both encoded using the common BLOSUM50 substitution matrix, with each residue represented by its corresponding row in the matrix. peptides with 8–15 amino acids long were converted as a 30-mer sequence by combing left and right-aligned representations, since our deep learning networks demand fixed-length inputs. The central gaps of peptides with less than 15 residues were filled with zero padding. For instance, "FLLVTLAIL" is represented by concatenating "FLLVTLAILXXXXXX" (left aligned), and "XXXXXXFLLVTLAIL" (right aligned), yielding the 30-mer sequence "FLLVTLAILXXXXXXXXXXXXLIALTVLLF". This design was primarily motivated by structural research on peptide-MHC complexes. Previous studies revealed that the termini of peptides often play a more critical role in binding than the central regions, as they are typically positioned in two specific binding pockets within the peptide binding grooves. Therefore, each ligand is encoded into a 30 × 20 matrix using such a pair-end approach.

HLA allele representation

The 34 amino acids derived from the multiple sequence alignment were used by the neural network to represent MHC-I molecules. According to the NetMHCpan tool, this representation is referred to a "pseudo-sequence". These residues were in close proximity to the peptide residues, specifically within 4.0 angstroms. The entire set of chosen locations were 31, 33, 48, 69, 83, 86, 87, 90, 91, 93, 94, 97, 98, 100, 101, 104, 105, 108, 119, 121, 123, 138, 140, 142, 167, 171, 174, 176, 180, 182, 183, 187, 191, 195, based on HLA-A*01:01 protein residue numbering (IMGT accession HLA00001), starting from 1. Using the BLOSUM50 substitution matrix, each amino acid was converted to a 36×20 matrix-vector, much as the ligand encoding.

Deep learning techniques

ImmuneApp implements a novel pan-allele MHC-I binding model that supports variable-length peptides of 8–15 amino acids. This model is developed using a hybrid deep learning architecture, which autonomously identifies critical residues and distinguishing features within the peptides. The architecture consists of five primary parts: a feature encoding module, a convolutional module, an LSTM module, an attention module, and an output layer. The model initially runs via the convolutional module (ConV) for initial feature extraction after passing the first feature encoding module through it:

$$Convol(L)_{ij} = \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} K_{rc}^M L_{i+r,c} \quad (2)$$

where L stands for the input antigen or MHC pseudo-sequence, i and j represent the indices for the output position and the kernel, respectively. K^M serves as a convolutional kernel by a $R \times C$ weight matrix, where R denotes the kernel's window size, and C represents the input dimension. To reduce the dimensionality of the MHC initial feature map, a max-pooling operator was implemented after the convolutional layer.

To effectively capture the intricate long-range relationships within the sequence, the model passes the extracted feature maps into an LSTM layer. The LSTM unit consists of four components: an input gate, a forget gate, an output gate, and a single cell that can recall characteristics across any time period. Specifically, taking a peptide with length T as input $\{x_p\}_{p=1}^{T}$ in LSTM, and for each position t, define the input gate as I_t, forget gate as F_t, output gate as O_t, hidden state as H_t and cell state as C_t. The steps in the LSTM training procedure are as follows:

$$F_t = \sigma\left(W_f \times [x_t, H_t - 1] + b_t\right) \quad (3)$$

$$I_t = \sigma\left(W_I \times [x_t, h_t - 1] + b_I\right) \quad (4)$$

$$C_t = F_t \times C_{t-1} - I_t \times \tanh\left(W_C \times [x_t, h_t - 1] + b_C\right) \quad (5)$$

$$O_t = \sigma\left(W_O \times [x_t, h_t - 1] + b_O\right) \quad (6)$$

$$H_t = O_t \times \tanh\left(C_t\right) \quad (7)$$

To learn all the hidden features within the LSTM layer and assign greater weight to critical locations, recurrent outputs are densely connected to an attention module. Mathematically, the attention mechanism generates an output vector by using the variables $\{B_t\}_{t=1}^{T}$ from LSTM layer. As demonstrated below:

$$\alpha_t = \frac{\exp(w(B_t))}{\sum_{i=1}^{T} \exp(w(B_i))} \quad (8)$$

$$As = \sum_{t=1}^{T} \alpha_t B_t \quad (9)$$

where w denotes a neural network calculating a scalar weight. A fully connected layer is formed by concatenating the outputs from both the LSTM and attention modules. The output layer applies a sigmoid nonlinear transformation to generate the probability of antigen presentation for specific HLA class I alleles.

Deconvolution of multi-allelic immunopeptidomics data

Using our curated MONOALLELIC training data, a new pan-binding prediction model (ImmuneApp-EL) was generated to estimate the likelihood that a query peptide is presented by an HLA-I allele. This training dataset encompassed 149 alleles and comprised 349,650 EL alongside 17,482,200 decoy peptides. To address the sample imbalance issue and enhance the robustness of the model, ImmuneApp-EL

was implemented using a balanced class-weight approach and the ensemble learning strategy. Specifically, "compute_class_weight" function (Python package sklearn) was applied to calculate class weights, which were added during the model fitting. Moreover, different batch sizes (2048, 4096, 8192, 16,384, and 32,768) were set. For each batch size, the MONOALLELIC-training dataset was divided in a 4:1 ratio, allocating data for training and validation. The number of training epochs was determined based on the maximization of AUPRC on the validation dataset, with a cap of 100 epochs imposed. The training process is repeated five times to ensure every subset is used for both training and testing (like fivefold cross-validation), resulting in the generation of 25 models. The final prediction score for each query was the average of the 25 models' outputs.

Using ImmuneApp-EL, we developed a model-driven deconvolution method to transform immunopeptidomics as pseudo-single-allelic ligands. For each sample, we initially made predictions for HLA alleles. To make the predicted scores for different alleles comparable in a sample set, we calibrated raw scores using percent rank values. To this end, percentileofscore function (Python package stats) was used to compute the percentile rank of a score for each allele relative to a list of scores in a background set of 500,000 decoy peptides. For each individual sample, we eliminated all allele-peptide pairs that had a predicted binding rank exceeding the 20% threshold, thereby discarding peptides that were unlikely to bind to any of the specified alleles. In instances where multiple alleles were predicted to bind with a particular peptide, the allele-peptide pair that had the lowest binding rank (indicating the strongest binder) was chosen.

Development of integrative antigen presentation model and immunogenicity predictor

Encouraged by the previous evidence that integrating multi-allelic ligands could improve the performance of antigen presentation prediction, we further processed and incorporated available immuno-peptidomics data to enhance model training. The final curated multi-allelic dataset comprises 969,435 ligands restricted to 110 HLA-I molecules from 216 samples. By employing our model-driven deconvolution method, we mapped 835,551 ligands to 104 alleles and obtained 328,227 unique HLA allele-ligand pairs. Subsequently, the mono-allelic and pseudo-mono-allelic datasets were merged. After the removal of duplicates, we compiled a total of 573,453 unique allele-ligand pairs representing 162 alleles (Supplementary Data 5). This comprehensive dataset was utilized as the final training set for a mixed prediction model, ImmuneApp-MA, following the aforementioned training strategy. In addition, accurate prediction of immunogenic neoepitopes, especially within the top-ranked outputs, helped in identifying potential targets for immunotherapeutic approaches, representing a challenge for most antigen-binding related predictors. Using the integrative antigen presentation model as a pre-trained model, we leveraged deep transfer learning into a curated dataset of immunogenicity to develop a new immunogenicity predictor. The training strategy involved fine-tuning the final three dense layers of the mixed prediction model using the immunogenicity dataset (ImmuneApp-Neo). We chose the five models with the best performance as the base model for transfer learning. In this study, neural networks are built with Keras 2.3 (https://keras.io/) and the Tensorflow backend in Python 3.7. To accelerate the gradient descent during training, we run on an NVIDIA Tesla T100 GPU server with CUDA 7.5 on our GPU clusters.

Comparison to existing methods

To further assess the performance of our models, we conducted several benchmarking analyses using external immunopeptidomics samples, encompassing both mono-allelic and multi-allelic datasets, as well as neoepitope immunogenicity data. We compared with seven methods: NetMHCpan-4.1[49], MHCflurry 2.0[51], MixMHCpred 2.1 and 2.2[50],

HLAthena[22], MHCnuggets-2.4[71], TransPHLA[72], and PRIME 1.0 and 2.0[50]. The PRIME tool was added only for the comparison of immunogenicity. These methods are well-established and widely used in the field. Both percentile rank outputs and prediction scores were used for comparative analysis. Three evaluation metrics, AUROC, AUPRC, and PPV, were calculated as follows:

$$\text{Sensitivity} = \frac{N(\text{correct predicted hits})}{N(\text{all hits})} \quad (10)$$

$$\text{Specificity} = \frac{N(\text{correct predicted decoys})}{N(\text{all decoys})} \quad (11)$$

$$\text{Recall} = \frac{N(\text{correct predicted hits})}{N(\text{correct predicted hits}) + N(\text{incorrect predicted decoys})} \quad (12)$$

$$\text{Precision} = \frac{N(\text{correct predicted hits})}{N(\text{correct predicted hits}) + N(\text{incorrect predicted hits})} \quad (13)$$

$$\text{PPV} = \frac{N(\text{correct predicted hits})}{N(\text{all hits predicted positive})} \quad (14)$$

N represents the total predicted results. AUROC scores were derived from the area under the curves representing sensitivity and $1 - $ specificity. AUPRC scores were determined from the area under the precision and recall curves. PPV highlighted the predictor's ability to prioritize true hits.

Benchmarking of antigen presentation prediction. The MONOALLELIC-testing dataset comprises 43,866 HLA-I ligands. This dataset was excluded from the training sets of all previous predictors so that it could provide an unbiased evaluation. A set of randomly selected peptides from the human proteome was utilized as negatives, with a 50-fold excess, to compute AUROC, AUPRC, and PPV for all predictors evaluated in this study (ImmuneApp-EL and ImmuneApp-MA) and other tools, including MixMHCpred 2.1&2.2, HLAthena, NetMHCpan-4.1, MHCflurry 2.0, MHCnuggets-2.4, and TransPHLA.

Benchmarking of immunopeptidomics deconvolution. To assess model performance with multi-allelic samples, 435,397 eluted ligands obtained from 47 recently published samples (the MULTIALLELIC-Recent benchmark) were used. These ligands were considered positives and were combined with a large number of randomly selected peptides. We evaluated our approach and other tools in two ways. First, AUROC, AUPRC, and PPV were calculated to assess the effectiveness of predictors in identifying true ligands within large random peptide libraries. Moreover, we examined the congruence between HLA binding motifs obtained through deconvolution and motifs identified by single-allelic ligands. Average PCC values among alleles were calculated. We assessed our methods in comparison with NetMHCpan4.1 and MixMHCpred 2.2, which employ NNalign-MA and MixMHCp for immunopeptidomics deconvolution, respectively.

Benchmarking of neoepitopes immunogenicity. To assess the potential clinical significance, we performed a comparative analysis of all predictors developed in the present study against peer tools for screening immunogenic neoepitopes. The dataset under evaluation included 349 immunogenic and 1838 non-immunogenic neoepitopes collected from diverse databases and studies (IMMUNOGENIC-testing data). It is crucial for antigen prediction methods to prioritize a significant proportion of immunogenicity in their top-ranking prediction

scores, as only a select few candidate neoantigens ranked at the top undergo clinical testing and practical application. Therefore, PPV was computed to evaluate the immunogenicity prediction for ImmuneApp-Neo, ImmuneApp-MA, ImmuneApp-EL, PRIME 1.0 and 2.0, HLAthena, NetMHCpan-4.1, MixMHCpred 2.1&2.2, MHCflurry 2.0, TransPHLA, and MHCnuggets-2.4.

Implementation of ImmuneApp online platform
ImmuneApp implements four main modules: "Discovery", "Analysis", "Results" and "Controller". In the backend, three well-trained deep learning models (ImmuneApp_BA, ImmuneApp_MA and ImmuneApp_Neo) are used for the predictions of binding affinities, ligand probabilities, and immunogenicity as well as immunopeptidomic analysis, respectively. The "Controller" module checks the input data format, sends the data from frontend interfaces to the backend, creates the results using models, and then provides the results on the "Results" page. The "Discovery" module accepts two input types: "FASTA" and "Peptide". Users can directly copy the input data to an online submission text box. Moreover, MHC molecules and the peptide length (only FASTA input) need to be specified for running prediction. The "Analysis" module accepts clinical immunopeptidomic samples as input, together with MHC molecules. The input sample(s) can be directly copied to an online submission text box or uploaded from the user's local disk. Sample identity should be specified. This module provides intuitive report for personalized analysis, statistical reports, and visualization of results for immunopeptidomic data. We implemented both pages in a responsive manner by using the HTML5, CSS, Bootstrap3, and JavaScript. Additionally, the "Controller" is called through Ajax technology to submit jobs, retrieve data, and show results. There is no limit to the number of tasks submitted by each user. ImmuneApp can automatically handle the jobs in a queue, which allows up to five jobs to execute concurrently.

Motif analysis and discovery for immunopeptidomics data
We implemented both unsupervised gibbscluster and supervised allele-specific approaches for motif analysis. The unsupervised GibbsCluster employs a standard GibbsCluster execution utilizing all available peptides. For this analysis, the parameters set were based on the recommended defaults for class I peptides provided by the GibbsCluster-2.0 server: "-g 1-6 -T -j 2 -C -D 4 -I 1". The grouping exhibiting the highest Kullback–Leibler distance (KLD) score will be detailed in the report. Additionally, the allele-specific approach relies on the outcomes of our model-driven deconvolution method, which transforms immunopeptidomics into pseudo-mono-allelic data for each allele. For peptides not predicted to bind to any allele, GibbsCluster was executed with the previously mentioned parameters and a range of "-g" values from 1–5. This approach enabled GibbsCluster to identify multiple groups within these unannotated peptides, with the grouping displaying the highest KLD score being highlighted.

Reporting summary
Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability
All datasets used in this study are publicly available and detailed in Supplementary Data 1-7. The datasets of curated mono-allelic HLA-I ligands are available in Supplementary Data 1 and Supplementary Data 2. The datasets of multi-allelic HLA-I immunopeptidomics are available in Supplementary Data 3 and Supplementary Data 4. The dataset of merged mono-allelic and pseudo-mono-allelic ligands used to train the composite model is included in Supplementary Data 5. The dataset of HLA-I neoepitopes is available in Supplementary Data 6. The dataset of MHC class I binding affinity is available in Supplementary Data 7. The sequences of different MHC molecules were obtained from

the IPD-IMGT/HLA Database (https://www.ebi.ac.uk/ipd/imgt/hla/). Relevant raw datasets for each figure are provided in the Source Data file. Source data are provided with this paper.

Code availability
Researchers can run ImmuneApp online at https://bioinfo.uth.edu/iapp/. For commercial usage inquiries, please contact the authors. The source codes are implemented in Python and are freely available at GitHub[73], with https://doi.org/10.5281/zenodo.13357725.

References
1. Iwasaki, A. & Medzhitov, R. Regulation of adaptive immunity by the innate immune system. *Science* **327**, 291–295 (2010).
2. Iwasaki, A. & Medzhitov, R. Control of adaptive immunity by the innate immune system. *Nat. Immunol.* **16**, 343–353 (2015).
3. Roy, P., Orecchioni, M. & Ley, K. How the immune system shapes atherosclerosis: roles of innate and adaptive immunity. *Nat. Rev. Immunol.* **22**, 251–265 (2022).
4. Neefjes, J., Jongsma, M. L., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**, 823–836 (2011).
5. Pishesha, N., Harmand, T. J. & Ploegh, H. L. A guide to antigen processing and presentation. *Nat. Rev. Immunol.* **22**, 751–764 (2022).
6. Lang, F., Schrörs, B., Löwer, M., Türeci, Ö. & Sahin, U. Identification of neoantigens for individualized therapeutic cancer vaccines. *Nat. Rev. Drug Discov.* **21**, 261–282 (2022).
7. Kilian, M. et al. MHC class II-restricted antigen presentation is required to prevent dysfunction of cytotoxic T cells by blood-borne myeloids in brain tumors. *Cancer Cell* **41**, 235–251.e239 (2023).
8. Harding, C. V. & Boom, W. H. Regulation of antigen presentation by Mycobacterium tuberculosis: a role for Toll-like receptors. *Nat. Rev. Microbiol.* **8**, 296–307 (2010).
9. Ott, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).
10. Balachandran, V. P. et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
11. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
12. Blass, E. & Ott, P. A. Advances in the development of personalized neoantigen-based therapeutic cancer vaccines. *Nat. Rev. Clin. Oncol.* **18**, 215–229 (2021).
13. Łuksza, M. et al. Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389–395 (2022).
14. Rosenthal, R. et al. Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479–485 (2019).
15. Xie, N. et al. Neoantigens: promising targets for cancer therapy. *Signal Transduct. Target. Ther.* **8**, 9 (2023).
16. Leidner, R. et al. Neoantigen T-cell receptor gene therapy in pancreatic cancer. *N. Engl. J. Med.* **386**, 2112–2119 (2022).
17. Yuan, M. et al. A highly conserved cryptic epitope in the receptor binding domains of SARS-CoV-2 and SARS-CoV. *Science* **368**, 630–633 (2020).
18. VanBlargan, L. A. et al. A potently neutralizing SARS-CoV-2 antibody inhibits variants of concern by utilizing unique binding residues in a highly conserved epitope. *Immunity* **54**, 2399–2416. e2396 (2021).
19. Shomuradova, A. S. et al. SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T cell receptors. *Immunity* **53**, 1245–1257.e1245 (2020).
20. Harndahl, M. et al. Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays. *J. Biomol. Screen.* **14**, 173–180 (2009).
21. Rasmussen, M. et al. Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule. *J. Immunol.* **193**, 4790–4802 (2014).

22. 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).

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

24. Bulik-Sullivan, B. et al. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nat. Biotechnol.* **37**, 55–63 (2019).

25. Sette, A. et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol. (Baltim., Md.: 1950)* **153**, 5586–5592 (1994).

26. Chong, C., Coukos, G. & Bassani-Sternberg, M. Identification of tumor antigens with immunopeptidomics. *Nat. Biotechnol.* **40**, 175–188 (2022).

27. Weingarten-Gabbay, S. et al. Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs. *Cell* **184**, 3962–3980.e3917 (2021).

28. Jaeger, A. M. et al. Deciphering the immunopeptidome in vivo reveals new tumour antigens. *Nature* **607**, 149–155 (2022).

29. Ouspenskaia, T. et al. Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.* **40**, 209–217 (2022).

30. Weingarten-Gabbay, S. et al. The HLA-II immunopeptidome of SARS-CoV-2. *Cell Rep.* **43**, 113596 (2024).

31. Huang, X. et al. The SysteMHC Atlas v2. 0, an updated resource for mass spectrometry-based immunopeptidomics. *Nucleic Acids Res.* **52**, D1062–D1071 (2024).

32. Kim, G. B. et al. Quantitative immunopeptidomics reveals a tumor stroma–specific target for T cell therapy. *Sci. Transl. Med.* **14**, eabo6135 (2022).

33. Ferreira, H. J. et al. Immunopeptidomics-based identification of naturally presented non-canonical circRNA-derived peptides. *Nat. Commun.* **15**, 2357 (2024).

34. Kina, E. et al. Breast cancer immunopeptidomes contain numerous shared tumor antigens. *J. Clin. Investig.* **134**, e166740 (2024).

35. 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).

36. Kacen, A. et al. Post-translational modifications reshape the antigenic landscape of the MHC I immunopeptidome in tumors. *Nat. Biotechnol.* **41**, 239–251 (2023).

37. Wen, B., Li, K., Zhang, Y. & Zhang, B. Cancer neoantigen prioritization through sensitive and reliable proteogenomics analysis. *Nat. Commun.* **11**, 1759 (2020).

38. Tretter, C. et al. Proteogenomic analysis reveals RNA as a source for tumor-agnostic neoantigen identification. *Nat. Commun.* **14**, 4632 (2023).

39. Martin, M. V. et al. The neo-open reading frame peptides that comprise the tumor framome are a rich source of neoantigens for cancer immunotherapy. *Cancer Immunol. Res.* OF1-OF20 (2024).

40. O'Donnell, T. J. et al. MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst.* **7**, 129–132.e124 (2018).

41. Hoof, I. et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**, 1–13 (2009).

42. Nielsen, M. & Andreatta, M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* **8**, 1–9 (2016).

43. Vita, R. et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43**, D405–D412 (2015).

44. Bassani-Sternberg, M. et al. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity. *PLoS Comput. Biol.* **13**, e1005725 (2017).

45. Bassani-Sternberg, M. & Gfeller, D. Unsupervised HLA peptidome deconvolution improves ligand prediction accuracy and predicts cooperative effects in peptide–HLA interactions. *J. Immunol.* **197**, 2492–2499 (2016).

46. 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).

47. Alvarez, B. et al. NNAlign_MA; MHC peptidome deconvolution for accurate MHC binding motif characterization and improved T-cell epitope predictions. *Mol. Cell. Proteom.* **18**, 2459–2477 (2019).

48. Chen, B. et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat. Biotechnol.* **37**, 1332–1343 (2019).

49. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**, W449–W454 (2020).

50. Gfeller, D. et al. Improved predictions of antigen presentation and TCR recognition with MixMHCpred2. 2 and PRIME2.0 reveal potent SARS-CoV-2 CD8⁺ T-cell epitopes. *Cell Syst.* **14**, 72–83.e75 (2023).

51. O'Donnell, T. J., Rubinsteyn, A. & Laserson, U. MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst.* **11**, 42–48.e47 (2020).

52. Racle, J. et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* **37**, 1283–1286 (2019).

53. Reynisson, B. et al. Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. *J. Proteome Res.* **19**, 2304–2315 (2020).

54. Wells, D. K. et al. Key parameters of tumor epitope immunogenicity revealed through a consortium approach improve neoantigen prediction. *Cell* **183**, 818–834.e813 (2020).

55. Schmidt, J. et al. Prediction of neo-epitope immunogenicity reveals TCR recognition determinants and provides insight into immunoediting. *Cell Rep. Med.* **2**, 100194 (2021).

56. Rapin, N., Hoof, I., Lund, O. & Nielsen, M. MHC motif viewer. *Immunogenetics* **60**, 759–765 (2008).

57. Tadros, D. M., Eggenschwiler, S., Racle, J. & Gfeller, D. The MHC Motif Atlas: a database of MHC binding specificities and ligands. *Nucleic Acids Res.* **51**, D428–D437 (2023).

58. Andreatta, M., Alvarez, B. & Nielsen, M. GibbsCluster: unsupervised clustering and alignment of peptide sequences. *Nucleic Acids Res.* **45**, W458–W463 (2017).

59. Abbott, C. W. et al. Prediction of immunotherapy response in melanoma through combined modeling of neoantigen burden and immune-related resistance mechanisms. *Clin. Cancer Res.* **27**, 4265–4276 (2021).

60. Shao, W. et al. The SysteMHC atlas project. *Nucleic Acids Res.* **46**, D1237–D1247 (2018).

61. Faridi, P. et al. A subset of HLA-I peptides are not genomically templated: Evidence for cis-and trans-spliced peptide ligands. *Sci. Immunol.* **3**, eaar3947 (2018).

62. Jappe, E. C. et al. Thermostability profiling of MHC-bound peptides: a new dimension in immunopeptidomics and aid for immunotherapy design. *Nat. Commun.* **11**, 6305 (2020).

63. Pyke, R. M. et al. Precision neoantigen discovery using large-scale immunopeptidomes and composite modeling of MHC peptide presentation. *Mol. Cell. Proteom.* **22**, 100506 (2023).

64. de Beijer, M. T. et al. Immunopeptidome of hepatocytes isolated from patients with HBV infection and hepatocellular carcinoma. *JHEP Rep.* **4**, 100576 (2022).

65. Qi, Y. A. et al. Proteogenomic analysis unveils the HLA class I-presented immunopeptidome in melanoma and EGFR-mutant lung adenocarcinoma. *Mol. Cell. Proteom.* **20**, 100136 (2021).

66. Pearson, H. et al. MHC class I–associated peptides derive from selective regions of the human genome. *J. Clin. Investig.* **126**, 4690–4701 (2016).
67. Xia, J. et al. NEPdb: a database of T-cell experimentally-validated neoantigens and pan-cancer predicted neoepitopes for cancer immunotherapy. *Front. Immunol.* **12**, 644637 (2021).
68. Kim, S. et al. Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. *Ann. Oncol.* **29**, 1030–1036 (2018).
69. Danilova, L. et al. The mutation-associated neoantigen functional expansion of specific T cells (MANAFEST) assay: a sensitive platform for monitoring antitumor immunity. *Cancer Immunol. Res.* **6**, 888–899 (2018).
70. Caushi, J. X. et al. Transcriptional programs of neoantigen-specific TIL in anti-PD-1-treated lung cancers. *Nature* **596**, 126–132 (2021).
71. Shao, X. M. et al. High-throughput prediction of MHC class I and II neoantigens with MHCnuggets. *Cancer Immunol. Res.* **8**, 396–408 (2020).
72. Chu, Y. et al. A transformer-based model to predict peptide–HLA class I binding and optimize mutated peptides for vaccine design. *Nat. Mach. Intell.* **4**, 300–311 (2022).
73. Xu, H. et al. ImmuneApp for HLA-I epitope prediction and immunopeptidome analysis. *ImmuneApp* https://github.com/bsml320/ImmuneApp, https://doi.org/10.5281/zenodo.13357725 (2024).

Author contributions

Z.Z .and H.X. conceptualized the project. H.X. and Z.L. designed the study and computational framework and analyzed the data and results. H.X. made figures and tables and implemented the model. R.H., L.K., and X.D. designed the website. H.X., R.H., Z.L., W.Z., C.T., and Z.Z. wrote and edited the paper. All authors have proofread and approved the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-53296-0.

Correspondence and requests for materials should be addressed to Haodong Xu, Zhihong Li or Zhongming Zhao.

Peer review information *Nature Communications* thanks Pouya Faridi, Josep Garnica and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.