



DeepHLApan: A Deep Learning Approach for Neoantigen Prediction Considering Both HLA-Peptide Binding and Immunogenicity

Jingcheng Wu^{1,2}, Wenzhe Wang², Jiucheng Zhang², Binbin Zhou², Wenyi Zhao^{1,2}, Zhixi Su³, Xun Gu⁴, Jian Wu², Zhan Zhou^{1*} and Shuqing Chen^{1*}

¹ Institute of Drug Metabolism and Pharmaceutical Analysis and Zhejiang Provincial Key Laboratory of Anti-Cancer Drug Research, College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China, ² College of Computer Science and Technology, Zhejiang University, Hangzhou, China, ³ MOE Key Laboratory of Contemporary Anthropology, School of Life Sciences, Fudan University, Shanghai, China, ⁴ Department of Genetics, Development and Cell Biology, Iowa State University, Ames, IA, United States

OPEN ACCESS

Edited by:

Irina Caminschi,
Monash University, Australia

Reviewed by:

Tim Elliott,
University of Southampton,
United Kingdom
Stephanie Gras,
Monash University, Australia

*Correspondence:

Zhan Zhou
zhanzhou@zju.edu.cn
Shuqing Chen
chenshuqing@zju.edu.cn

Specialty section:

This article was submitted to
Antigen Presenting Cell Biology,
a section of the journal
Frontiers in Immunology

Received: 31 May 2019

Accepted: 15 October 2019

Published: 01 November 2019

Citation:

Wu J, Wang W, Zhang J, Zhou B,
Zhao W, Su Z, Gu X, Wu J, Zhou Z
and Chen S (2019) DeepHLApan: A
Deep Learning Approach for
Neoantigen Prediction Considering
Both HLA-Peptide Binding and
Immunogenicity.
Front. Immunol. 10:2559.
doi: 10.3389/fimmu.2019.02559

Neoantigens play important roles in cancer immunotherapy. Current methods used for neoantigen prediction focus on the binding between human leukocyte antigens (HLAs) and peptides, which is insufficient for high-confidence neoantigen prediction. In this study, we apply deep learning techniques to predict neoantigens considering both the possibility of HLA-peptide binding (binding model) and the potential immunogenicity (immunogenicity model) of the peptide-HLA complex (pHLA). The binding model achieves comparable performance with other well-acknowledged tools on the latest Immune Epitope Database (IEDB) benchmark datasets and an independent mass spectrometry (MS) dataset. The immunogenicity model could significantly improve the prediction precision of neoantigens. The further application of our method to the mutations with pre-existing T-cell responses indicating its feasibility in clinical application. DeepHLApan is freely available at <https://github.com/jjujiezz/deephlapan> and <http://biopharm.zju.edu.cn/deephlapan>.

Keywords: deep learning, neoantigen, recurrent neural network, human leukocyte antigen, cancer immunology

INTRODUCTION

Cancer cells are different, through somatic mutations, from normal cells and can therefore be recognized being a foreign cell by the immune system. Among all foreign elements in cancer cells, neoantigens are the most widely recognized elements derived from mutated genes (1). Neoantigens have therefore been acknowledged as ideal targets for cancer immunotherapies, such as cancer vaccines and T-cell immunotherapies (2–5). Recent studies also indicated that neoantigens are closely related to the therapeutic effect of immune checkpoint blockade therapies (6–8). However, only a small number of somatic mutations can generate neoantigens. It is still challenging to identify somatic mutations that can generate effective neoantigens (9).

Recently, whole exome sequencing combined with bioinformatic prediction has been widely used for candidate neoantigen identification (10). Several computational pipelines, such as TSNAD (11), pVAC-Seq (12), and INTEGRATE-neo (13), have been developed for this purpose. The most critical component of these pipelines is the *in-silico* estimation of binding between human leukocyte

antigens (HLAs) and peptides. Previous prediction methods on peptides binding with HLA alleles can be categorized into three groups, including (i) position-specific scoring matrix (PSSM)-based methods (14, 15), (ii) machine learning-based methods (16, 17), and (iii) structure-based methods (18, 19). There are also some consensus methods that combine several methods for better predictive performance (20, 21). Several neoantigen databases have also been developed based on pan-cancer immunogenomic analyses using neoantigen prediction tools, such as TSNAdb (22) and TCIA (23). However, existing tools are insufficient for neoantigen prediction in clinical applications because few of the predicted binders are immunogenic (1). Many attempts have been made to improve the prediction accuracy. For instance, several researchers have performed epitope prediction based on mass spectrometry (MS) profiling of HLA-peptide sequences. This method also considered proteasomal cleavage and transporter-associated with antigen processing (TAP)-mediated peptide transport, which are necessary in antigen presentation (24–27). In addition, the developers of NetMHCpan have released their latest version, which is trained based on both binding affinities and MS data of HLA-peptide binding (17). Moreover, the rapid development of deep learning methods, such as Convolutional Neural Networks (CNNs), has led to an increase in their use in cancer immunology (28–32). However, current methods will never catch up to the need for clinical applications unless they consider the potential immunogenicity of presented mutant peptides complexed with HLA class I molecules (pHLA). Several studies have considered the binding of pHLA and T-cell receptors (TCRs) to select immunogenic neoantigens using a neoantigen immune fitness model, which assumes that the TCR recognition probability of pHLA is positively correlated with the similarity between mutant peptides and pathogenic antigens (33–35). Furthermore, Jurtz et al. trained a sequence-based predictor of the interaction between TCRs and peptides presented by HLA-A02:01, which considers the sequences of both TCRs and peptides (36).

In this study, we develop a novel Recurrent Neural Network (RNN)-based approach, named DeepHLApan, for neoantigen prediction, considering both the binding between HLA-peptide pairs and the potential immunogenicity of pHLA. DeepHLApan consists of two models: the binding model for predicting the probability of the peptide being presented to the tumor cell membrane by HLA and the immunogenicity model for predicting the potential of pHLA eliciting T-cell activation. Since there are much more binding data than immunogenicity data, we take the prediction score of the immunogenicity model as a filter and rank the prediction scores of the binding model for high-confidence neoantigen identification. The binding model of DeepHLApan achieves comparable performance with other well-known binding prediction tools on the latest Immune Epitope Database (IEDB) benchmark and an independent MS dataset. We also confirm that the immunogenicity model could significantly improve the performance of other neoantigen prediction tools from previously published work (improvement ranging from 20.5 to 55.4%). We further applied our method to mutations with pre-existing T-cell responses and ranked most of them (69%) in the top 20, under an expression threshold

of transcripts per million (TPM) >2. These results indicate that DeepHLApan, combining the binding model and the immunogenicity model, could be beneficial for high-confidence neoantigen prediction and could further contribute to tumor immunotherapy in practice.

MATERIALS AND METHODS

HLA-Peptide Binding and Immunogenicity Data

The binding data between HLA class I alleles and peptides were collected from the IEDB (<http://www.iedb.org/>) (37). HLA-peptide pairs were filtered using the following criteria: (1) HLA class I alleles that are of HLA-A, B, and C subtypes, (2) the length of peptides range from 8 to 15, and (3) the pairs with inconsistent experimental results are excluded. In total, we obtained 327,178 non-redundant HLA-peptide pairs covering 169 HLA alleles with 257,089 of them being binders (**Figure S1A**, **Table S1**). We then balanced the positive data and negative data of each allele by following these steps: (1) train a basic model based on the 327,178 pairs, (2) create pseudo-HLA-peptide pairs. The number of created pseudo pairs of each HLA allele N_{pse} was calculated as follows:

$$N_{pse} = 100 * |N_{pos} - N_{neg}| \quad (1)$$

where N_{pos} and N_{neg} represent the number of positive and negative pairs of each HLA allele, respectively. Each peptide is generated by selecting one protein in the Ensembl database (38) randomly and extracting 8–11 mer peptides randomly in this protein without mutations, and (3) predict the binding possibility of pseudo pairs using the basic model and selecting the high-confidence negative (score <0.1) or positive (score >0.9) pairs for each allele.

After these steps, most of the alleles had balanced data, and the alleles with unusual proportions ($\frac{N_{pos}}{N_{neg}} > 5$ or $\frac{N_{neg}}{N_{pos}} > 5$) were removed (**Figure S2**, **Table S2**). Finally, 437,077 HLA-pairs covering 81 HLA alleles were used for training the binding model, with 280,525 collected pairs and 156,552 pseudo pairs (**Figure S1B**, **Table S3**).

The immunogenicity data of pHLA were also retrieved from the IEDB (37). The criterion for data filtering is as follows: (1) The length of peptides range from 8 to 15, and (2) for pairs with inconsistent experimental results, we selected the positive pairs. Finally, we obtained 32,785 HLA-peptide pairs with 5,702 of them related to HLA-A02:01 (**Table S4**). Among the 32,785 HLA-peptide pairs, 7,212 of them are immunogenic and 3,013 immunogenic HLA-peptide pairs are related to HLA-A02:01.

IEDB Benchmark Data

The data used to test the performance of the binding model were derived from the IEDB weekly benchmarking website (http://tools.iedb.org/auto_bench/mhci) (version 2018-05-11). It should be noted that this benchmark dataset includes 14 sub-datasets, but three of them with SLA molecules from *Sus scrofa* were not suitable for the binding model and were thus excluded. We then renumbered the selected 11 sub-datasets for convenience, and

the datasets measured by binding affinity were transformed to the binary type under the threshold as IEDB stated: HLA-A03:01 is 602 nM, HLA-A02:01 is 255 nM, HLA-B07:02 is 687 nM, and HLA-B27:05 is 584 nM. To note, all the data in this benchmark dataset were not included in previous training data. Meanwhile, the prediction performance of 12 models on these datasets was downloaded from the IEDB for comparison.

Independent MS Dataset

We collected an independent dataset from Mei et al. (39) for further binding model evaluation. The dataset has never been used for training in all previous developed tools as the authors claimed, but some of them have been used in the training of the binding model of DeepHLApan. We removed the sub-datasets that have been used for training, 15 sub-datasets (all of them are MS data) were retained and used for the model comparison.

Neoantigen Dataset Used for Evaluating Immunogenicity Model

The data used for evaluating the immunogenicity model were collected from Koşaloğlu-Yalçın et al. (40), containing 64 neoantigens with 6,400 random peptides that were generated based on mutation data extracted from The Cancer Genome Atlas (TCGA) database. In their work, NetMHCpan achieved excellent performance based on the AUC under rank thresholds of 10 and 2%.

CD8⁺ T-Cell Epitopes

We collected 2,023 assayed single-nucleotide variants from 17 patients, including 26 mutations with pre-existing T-cell responses from Bulik-Sullivan et al. (27), which combined the data from four published works (41–44) and substituted RNA-Seq data from tumor-type-matched patients of TCGA.

For the mutations from Tran et al. (41), Gros et al. (42), and Zacharakis et al. (44), all 8–11 mer peptides covering the mutations were extracted for prediction and resulted in 59,726 peptides and 372,252 HLA-peptide pairs (Table S5A). Normally, there would be 8 8-mer, 9 9-mer, 10 10-mer, and 11 11-mer peptides for each mutation; however, some of the mutations are located at the start or end of proteins and possess fewer than 38 peptides. For the mutations from Strønen et al. (43), the provided 2,852 HLA-peptide pairs were used for prediction (Table S5B). To evaluate the possibility of a mutation eliciting T-cell activation predicted by DeepHLApan, we removed the HLA-peptide pairs with predicted immunogenic scores <0.5. We then summed the binding scores of the rest of the pairs with the following formula for a mutation rank within one patient.

$$\Pr(\text{mutation}) = \sum_{i=8,9,10,11} \frac{\sum BS_{i\text{mer},h} * r_{i\text{mer},h}}{n_{i\text{mer}}} \quad (2)$$

where $\Pr(\text{mutation})$ is the probability of the mutation presentation, $BS_{i\text{mer},h}$ is the predicted binding scores of the i -mer peptides with HLA h , $r_{i\text{mer},h}$ is the actual ratio of i -mer peptides binding with HLA h in the training dataset, and $n_{i\text{mer}}$ is the number of i -mer within one mutation.

The mutation rank data of EDGE and MHCflurry were derived from Bulik-Sullivan et al. (27). The mutation rank of NetMHCpan 4.0 was measured by taking the minimum predicted rank across all mutation-spanning peptides.

We also collected 31 validated immunogenic HLA-peptide pairs from Tran et al. (41), Gros et al. (42), and Strønen et al. (43), which were not provided by Bulik-Sullivan et al. (27) (Table S6). Here, we separated the minimal epitopes according to their length and ranked them. Pairs with predicted immunogenic scores <0.5 were ignored, and the remaining pairs were ranked by the binding score.

Recurrent Neural Networks

The model architecture used for training was stacked by three layers of bidirectional Gated Recurrent Unit (BiGRU) with an attention layer. A Gated Recurrent Unit (GRU) is a variant of RNN which was first proposed by Cho et al. (45). Similar to the RNN, GRU handles the variable-length sequence by having a recurrent hidden state whose activation at each time is dependent on that of the previous time. The difference between GRU and RNN is the update of the recurrent hidden state which is the core part for overcoming gradients vanish or explode in training model to capture long-term dependencies. In particular, GRU proposes to derive the vector representations of hidden states h_t for each time step t as follows:

$$\begin{aligned} z_t &= \sigma(W_z x_t + U_z h_{t-1} + b_z) \\ r_t &= \sigma(W_r x_t + U_r h_{t-1} + b_r) \\ h_t &= (1 - z_t) \circ h_{t-1} + z_t \circ \phi(W_h x_t + U_h (r_t \circ h_{t-1}) + b_h) \end{aligned} \quad (3)$$

where x_t is the input vector, h_t is the output vector, z_t is the update gate vector, r_t is the reset gate vector, W , U and b are parameter matrices and vector, σ is the logistic sigmoid function, and ϕ is a hyperbolic tangent.

As for the BiGRU, it splits the neurons of a GRU into two directions, one for positive time direction (forward states), and another for negative time direction (backward states). By using two time directions, input information from the past and future of the current time frame can be used.

Attention Module

The attention module was incorporated with the BiGRU module to make modeling of long-term dependencies of 49 amino acids easier. The attention was first proposed by Bahdanau et al. (46) but we used the type of attention proposed by Raffel and Ellis (47). Given a model which produces a hidden state h_t at each time step, attention-based models compute a “context” vector c as the weighted mean of the state sequence h by

$$e_t = \sigma_a(h_t), \alpha_t = \frac{\exp(e_t)}{\sum_{k=1}^T \exp(e_k)}, c = \sum_{t=1}^T \alpha_t h_t. \quad (4)$$

where σ_a is a learnable function which only depends on h_t .

Model Training

Before training, the HLA alleles were transformed into pseudo-sequences as presented by the NetMHCpan (48) method for pan-allele prediction (i.e., each HLA allele was transformed into 34

amino acid residues located within 4.0 Å of the peptide). Then, the peptides were concatenated with the HLA pseudo-sequences. If the length of the combined sequence was <49 amino acids, pseudo amino acid “X” would be used for padding. The one-hot method was used for amino acid representation (i.e., transform each amino acid into a unique vector of 20 zeros and one 1).

Two steps were applied for reliable model training. First, the original dataset was used for basic model training. Then, the preliminary model was used for selecting high-confidence pseudo positive/negative HLA-peptide pairs. The selected pseudo pairs were added to the original dataset to balance the training data for training the final model. Other parameters are as follows: dropout rate was set to 0.2, the sigmoid function was used as the activation function, Binary Cross Entropy (BCE) was employed for loss calculation, and an Adam optimizer with a default learning rate of 0.001 was used for parameter optimization.

5-Fold Cross-Validation

Five-fold cross-validation is used to evaluate model robustness. Before training, the dataset is randomly partitioned into five non-overlapping subsets. The cross-validation process is repeated five times, with each subset used as a validation set while the remaining subsets are used as the training set. The results of the five validation sets are averaged to obtain the final result. One hundred epochs are executed, and the model is saved if the validation accuracy is better than previous epochs.

Evaluation Indicators

The area under the receiver operating characteristic curve (AUC) is the main measurement for model and software comparison. Accuracy (ACC) is used for the performance evaluation of the binding model on single-labeled HLA alleles. Precision [$\text{true positives}/(\text{true positives} + \text{false positives})$] and recall [$\text{true positives}/(\text{true positives} + \text{false negatives})$] are used to illustrate the importance of the immunogenicity model.

Data and Software Availability

All datasets for this study are included in the manuscript and the **Supplementary Files**. The source codes of DeepHLApan are freely available at <https://github.com/jiujiezz/deephapan>. The web service of DeepHLApan is available at <http://biopharm.zju.edu.cn/deephapan>.

RESULTS

The Architecture of DeepHLApan

Binding affinity is the most widely used measurement for HLA-peptide binding. However, a large number of HLA-peptide binding data without binding affinity have emerged with the development of high-throughput methods (e.g., MS) for HLA-peptide binding detection. In the binding dataset collected from the IEDB, 66.8% of the binding data were not measured by binding affinity (**Table S1**). And the gap further increases over time. Therefore, we transformed the binding data with binding affinity into a binary model to create a model that only provides the result of binding or not-binding. In addition, the potential immunogenicity of presented pHLA is a necessary factor for

tumor immunotherapy. However, published tools rarely consider the immunogenicity of pHLA due to insufficient data. We attempted to determine whether existing immunogenicity data could promote the identification of high-confidence neoantigens. DeepHLApan contains two independent parts (the binding model and the immunogenicity model), and its model framework consists of three layers of BiGRU with attention (**Figure 1**).

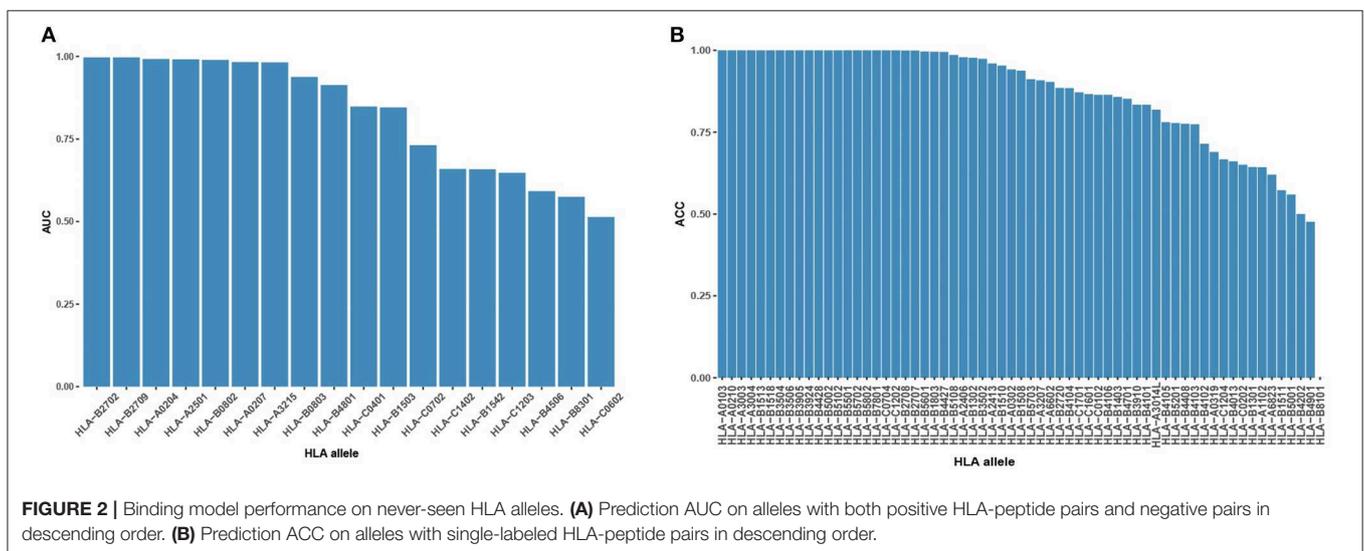
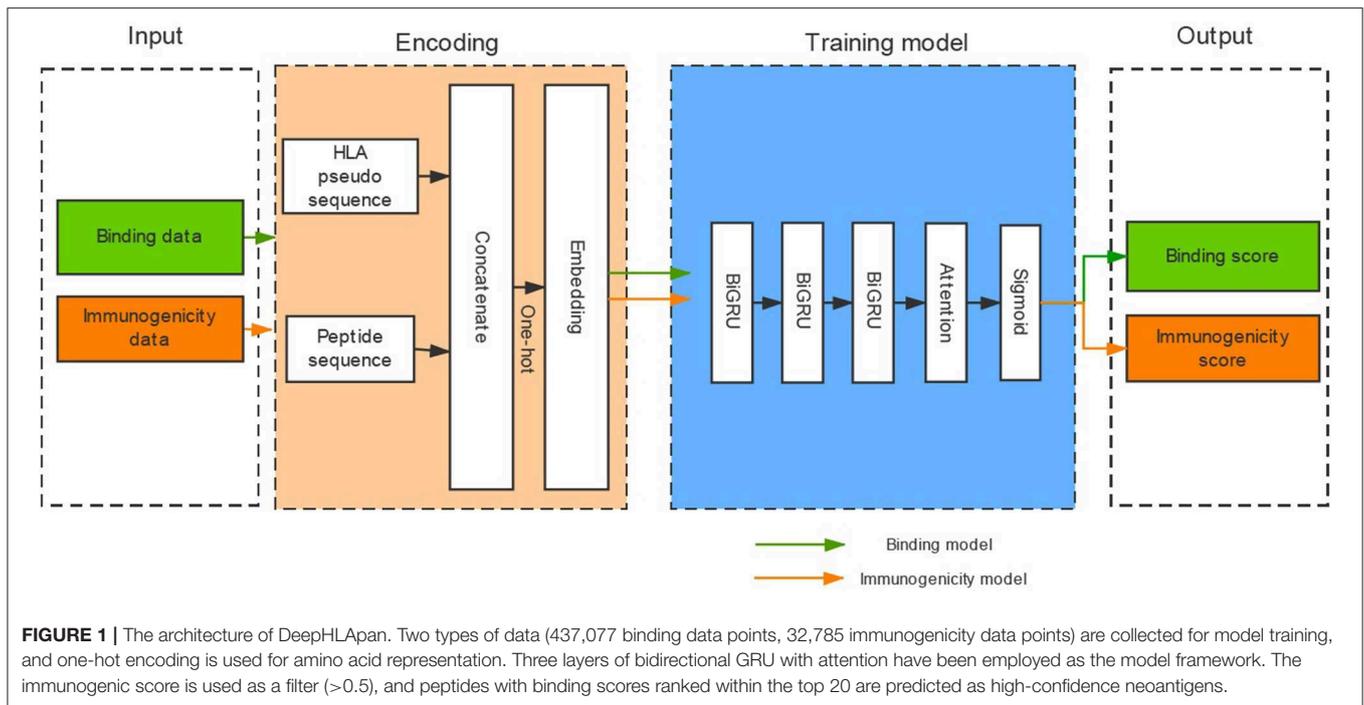
The Prediction Performance of the Binding Model on Unseen Alleles

During the binding model training, we created high-confidence pseudo HLA-peptide pairs to balance the training data of each allele and removed the HLA alleles with an imbalanced number of positive/negative pairs (see Materials and Methods). Of these removed alleles, 62 had pure positive HLA-peptide pairs, one had pure negative pairs, 14 had positive pairs more than 5-fold negative pairs and four had negative pairs of more than 5-fold positive pairs (**Figure S2**). We evaluated the performance of the binding model on these never-before-seen HLA alleles to discover its ability for pan-allele prediction. For the 18 alleles with both negative and positive pairs, most (9 out of 18, 50%) of them had AUC >0.9 (**Figure 2A**). For the rest of the untrained alleles, the majority (35 out of 63, 56%) had an accuracy >0.9 (**Figure 2B**). The reliable prediction results of the binding model on never-seen HLA alleles indicate that it is suitable for pan-allele prediction.

The binding model achieved good performance (AUC or accuracy >0.9) on 43 HLA alleles. To further investigate whether the binding model is able to learn truly correct motifs (only considering 9-mer) for these alleles, we compared the actual motifs and predicted motifs on 16 HLA alleles with more than 100 binding peptides (selected from 43 HLA alleles). The actual motifs were based on the binding peptides and the predicted motifs were generated by taking the top 1% predicted peptides out of 100,000 random peptides. WebLogo (49) was used for motif representation. We found that the predicted motifs are not exactly the same as the actual motifs, but they have similar patterns in most of the HLA alleles (**Figure 3**). The predicted motif of HLA-B48:01 is most dissimilar to the actual motif, but the amino acid arginine, glutamate, and glutamine are also conserved in the second amino acid residue of the predicted motif. These results demonstrate that the binding model of DeepHLApan does not simply boil down to a mix between many alleles of the training set, but has the ability to distinguish different alleles.

The Binding Model Achieves Comparable Performance With Other HLA-Peptide Binding Prediction Tools

The IEDB benchmark dataset is often used to compare the performance of different binding prediction tools. We used the latest IEDB benchmark dataset for this purpose. In the 11 sub-datasets that are suitable for all tools, the DeepHLApan binding model achieves the best performance for 6 out of the 11 (54.5%) sub-datasets while none of the other tools achieve

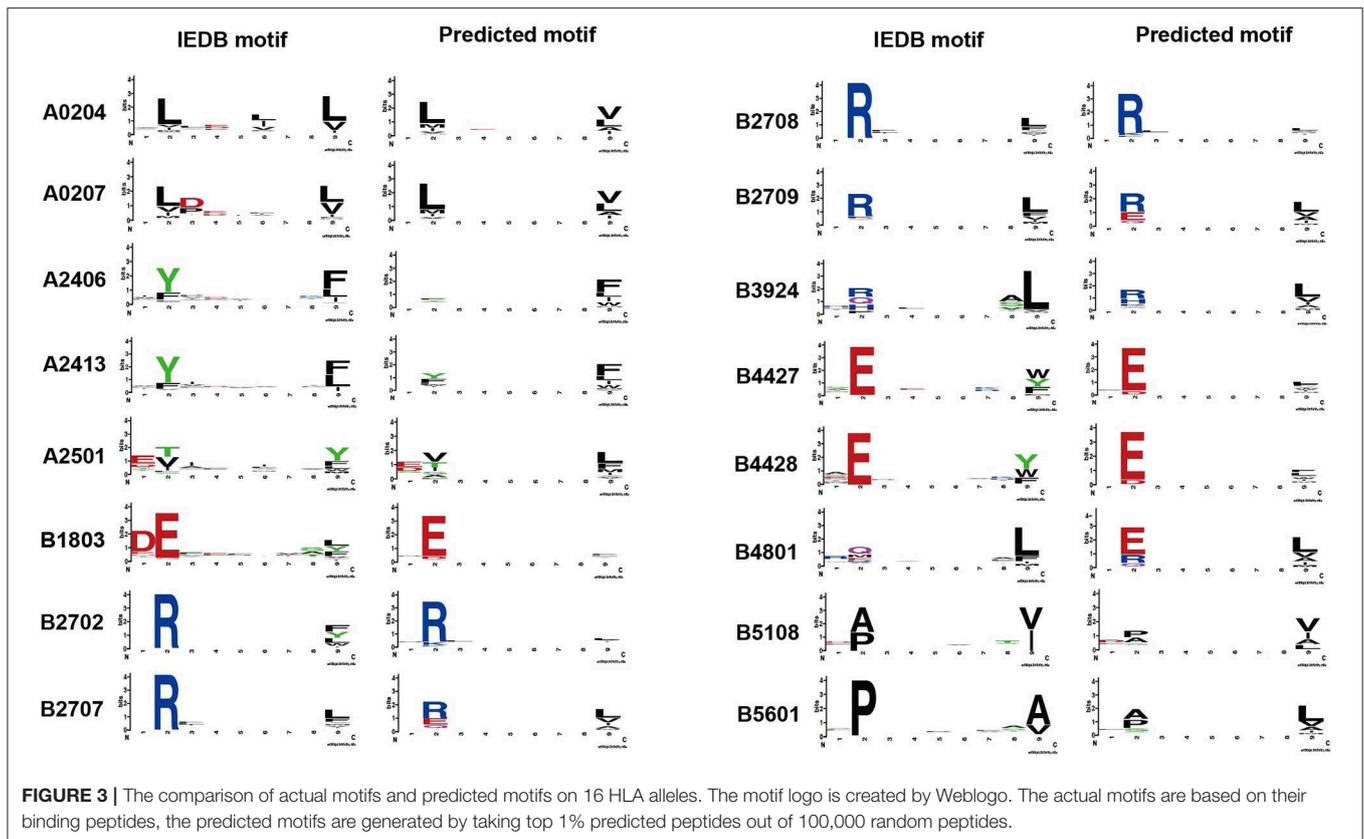


a best performance in more than four sub-datasets (**Figure 4A**, **Table S7A**).

Interestingly, all the tools have a similar prediction ability in nearly all the sub-datasets. For instance, in sub-datasets 1, 3, 4, 10, and 11, all the tools achieved good performance, with some of them having an AUC of 1, which could be due to having enough corresponding HLA allele-specific HLA-peptide pairs in the training data. However, none of the tools could obtain excellent performance in sub-datasets 5, 6, 7, 8, and 9. Further investigation found that these datasets were derived from Liepe et al. (50), and one third of them contained binding information between spliced peptides and HLA alleles. Because all the tools

were trained exclusively on natural (non-spliced) epitopes, the intrinsic differences between spliced and non-spliced antigenic peptides resulted in their limited prediction performance on spliced peptides (50).

We also collected an independent dataset from Mei et al. (39) for binding model evaluation (see Materials and Methods). Mei et al. provided a comprehensive analysis and benchmarking of 15 currently available tools for HLA-I peptide-binding prediction. The validation dataset they used has never been used for training in all previous developed tools and that's why only 5 tools [Pickpocket (51), MixMHCpred (25), NetMHCpan 4.0 (17), NetMHCcons (21), and NetMHCstabpan (52)] could obtain the



AUC values in all datasets. After removing the sub-datasets that have been used for training in the binding model of DeepHLApan, 15 sub-datasets were retained and had been used for model comparison. The results showed that all tools achieved similar performance on these datasets while MixMHCpred performed slightly better than others (**Figure 4B**, **Table S7B**).

The methods focusing on HLA-peptide binding have been developed over many years. Currently available methods, such as MixMHCpred and NetMHCpan 4.0, have achieved excellent performance on the basis of available experimental data. Based on the performance comparison on the IEDB benchmarking dataset and the independent MS dataset, we concluded that the DeepHLApan binding model could obtain comparable performance with state-of-the-art HLA-peptide binding tools.

The Immunogenicity Model Improves the Precision of Neoantigen Prediction Significantly

Because of the scarcity of immunogenicity data, few tools focus on the potential immunogenicity of pHLA, resulting in a high false positive rate (FPR) in practical applications. In this section, we explore whether existing immunogenicity data can contribute to neoantigen identification. Data used for immunogenicity model training were retrieved from the IEDB, and the validated neoantigens were retrieved from Koşaloğlu-Yalçın et al. (40) (see Materials and Methods). In their study, NetMHCpan obtained remarkably good performance in terms of the AUC. However, the predicted precision on their datasets was 9.6 and 36.6%

under the thresholds of 10 and 2%, respectively (**Table 1**). We added the predicted score from the immunogenicity model as an additional filter for neoantigen identification. The results showed that the predicted immunogenic score could improve the precision significantly (43.8 and 32.8% improvement under the thresholds of 10 and 2%, respectively) at the cost of less recall (**Table 1**, **Table S8**), which is acceptable for the purpose of predicting more reliable neoantigens rather than obtaining all potential neoantigens in practice. In addition, we also retrained an HLA-A02:01-restricted immunogenicity model to determine if more training data for each allele would attain both high precision and high recall due to the sufficient immunogenicity data of HLA-A02:01. The results showed that the HLA-A02:01-restricted model could improve the precision significantly (55.4 and 20.5% improvement under thresholds of 10 and 2%, respectively) and retain the recall with a decrease of <10% (**Table 2**, **Table S9**), indicating that the immunogenicity model could greatly contribute to high-confidence neoantigen identification with a growing amount of training data.

Application of DeepHLApan on Published CD8⁺ T-Cell Epitopes

In recently published work (27), the EDGE model, which was trained by a tumor HLA-peptide MS dataset, achieved better performance on retrospective neoantigen T-cell data than MHCflurry (53), a tool for binding affinity prediction. To directly evaluate the performance of DeepHLApan on neoantigen prediction, we applied it on the same data they

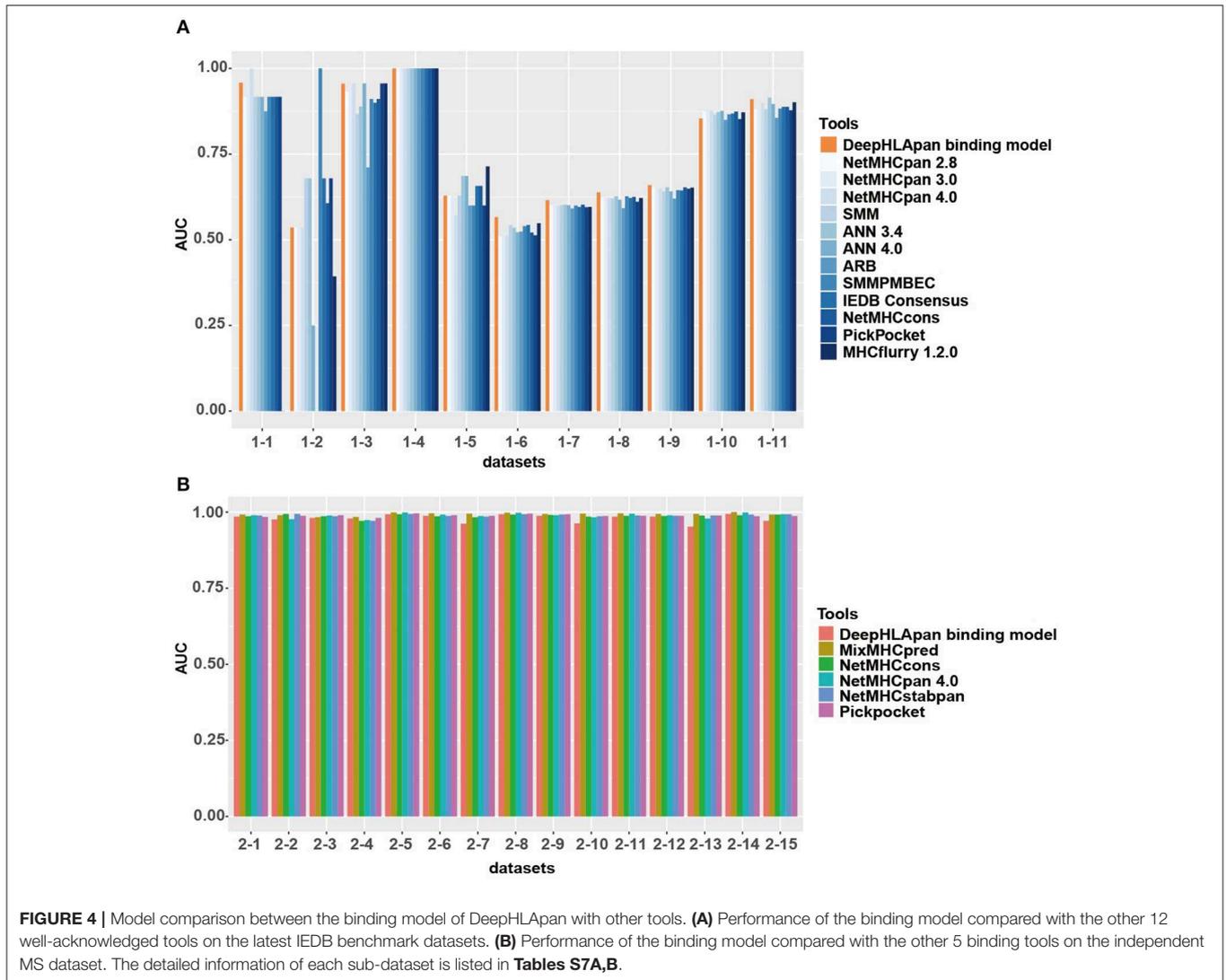


FIGURE 4 | Model comparison between the binding model of DeepHLApan with other tools. **(A)** Performance of the binding model compared with the other 12 well-acknowledged tools on the latest IEDB benchmark datasets. **(B)** Performance of the binding model compared with the other 5 binding tools on the independent MS dataset. The detailed information of each sub-dataset is listed in **Tables S7A,B**.

TABLE 1 | The improvement of precision and decrease of recall with immunogenicity model on all available neoantigen predictions.

Threshold (%)	NetMHCpan		NetMHCpan with immunogenicity model	
	Precision (%)	Recall (%)	Precision	Recall
10	9.6	96	13.8% (+43.8%)	56.3% (-41.9%)
2	36.6	92.2	48.6% (+32.8%)	54.7% (-40.7%)

The number in parentheses with a positive or negative sign indicates the percentage of improvement and decrease, respectively.

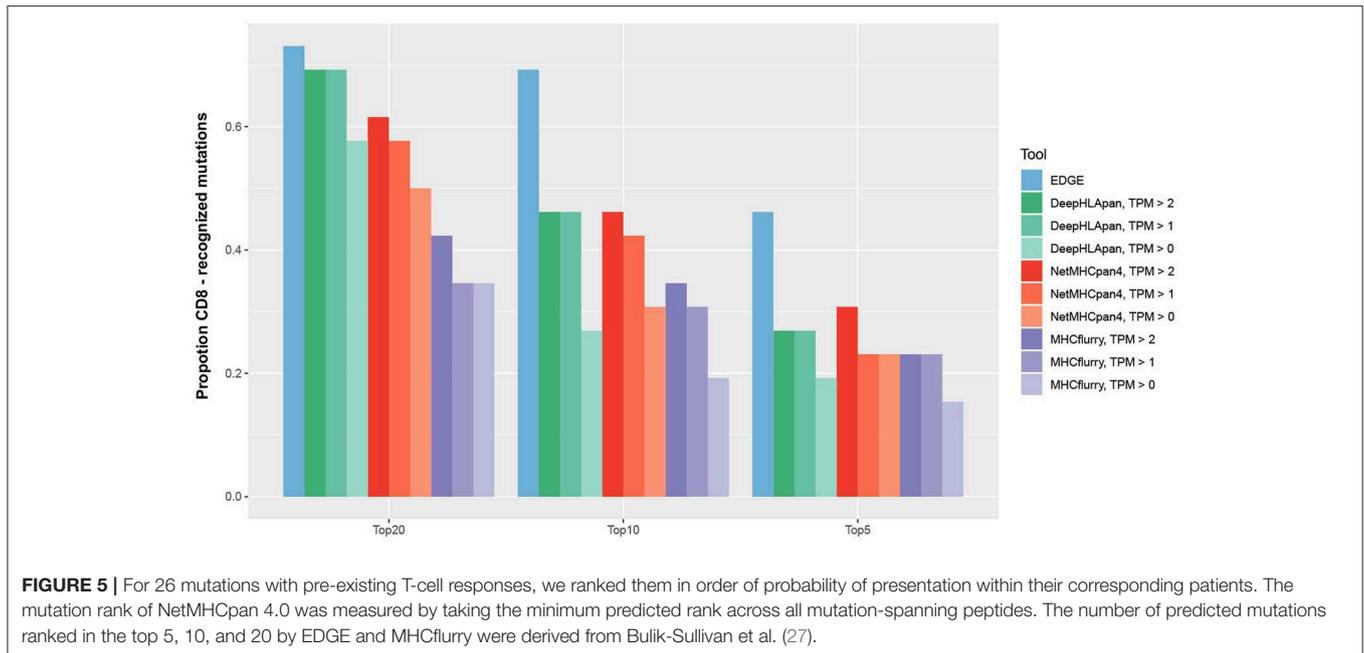
TABLE 2 | The improvement of precision and decrease of recall with immunogenicity model on HLA-A02:01-restricted neoantigen prediction.

Threshold (%)	A02:01 restricted		A02:01 restricted with immunogenicity model	
	Precision (%)	Recall (%)	Precision	Recall
10	10.1	96.0	15.7% (+55.4%)	88.0% (-8.3%)
2	37.5	84.0	45.2% (+20.5%)	76.0% (-9.5%)

The number in parentheses with a positive or negative sign indicates the percentage of improvement and decrease, respectively.

collected from four published works (41–44) (see Materials and Methods), which contained 2,023 mutations from 17 patients, 26 of which had pre-existing T-cell responses. We compared the performance of DeepHLApan with EDGE, MHCflurry and NetMHCpan 4.0 at the mutation level (**Table S10**). For each mutation, all 8–11 mer overlapping mutations were used for prediction and we ranked the mutations for each tool as described

in the Materials and Methods. Taking the number of pre-existing T-cell responses in the 5, 10, or 20 top-ranked mutations for each patient as the measurement, DeepHLApan performed better than MHCflurry, but is comparable with NetMHCpan 4.0 for the 5 and 10 top-ranked mutations across different gene expression thresholds. With the 20 top-ranked mutations, DeepHLApan achieved better performance than NetMHCpan



4.0 and a comparable performance with EDGE at TPM >2 (Figure 5). Further evaluation of the AUC performance of four tools on each patient showed that EDGE performed best and DeepHLApan performed better than the other two tools on average under different thresholds of TPM (Table S11). We also evaluated the performance of DeepHLApan on the rank of the concrete HLA-peptide pairs within one patient. DeepHLApan ranked 32.2% (10/31) of the immunogenic HLA-peptide pairs as the 20 top-ranked HLA-peptide pairs at TPM >2 with 32.2% (10/31) missed due to predicted immunogenic scores <0.5 (Table S6). The rate could be 47.6% (10/21) if we ignored the missed HLA-peptide pairs and is better than that predicted by EDGE (27). All of the results mentioned above indicated that DeepHLApan could identify high-confidence neoantigens by filtering with an immunogenic score >0.5 and selecting HLA-peptide pairs with 20 top-ranked binding scores.

DISCUSSION

Neoantigens have been acknowledged as ideal targets for tumor immunogenicity and substantial effort has been made in neoantigen identification. However, most of the existing tools only consider the binding affinity between human leukocyte antigens (HLAs) and peptides and achieve unsatisfactory results. Recently, the EDGE model, which was trained by a tumor HLA peptide mass spectrometry dataset, achieved excellent performance in selecting high-confidence neoantigens (27). However, it did not consider the potential immunogenicity of predicted pHLA, as did many other tools, which cannot be ignored in the process of T-cell activation.

In this study, we propose a novel RNN-based method, DeepHLApan, for high-confidence neoantigen prediction considering both the possibility of mutant peptide presentation and the potential immunogenicity of pHLA. We demonstrate that the binding model could achieve good performance

on unseen HLA alleles and has a comparable performance with other well-acknowledged tools on the latest IEDB benchmark datasets and an independent MS dataset. In the model comparison of the DeepHLApan binding model with other tools trained on datasets of canonical peptides, all of them performed poorly on the datasets with a high number of spliced peptides. One possible reason might be the different binding patterns between spliced and non-spliced antigenic peptides. Another reason might be that the identification of many of the spliced peptides stated in Liepe et al. (50) are ambiguous, and the vast majority of these peptides likely correspond to false-positives (54).

Using the immunogenicity model on the neoantigen datasets collected from Koşaloğlu-Yalçın et al. we demonstrate that the predicted immunogenic score could significantly improve prediction precision of neoantigens. Although, with the improvement of precision of neoantigen identification by the immunogenicity model, the recall rate decreases in the neoantigen prediction, the cost is acceptable in clinical applications because the main problem in neoantigen identification is the high FPR for tumors with a lot of mutations. The decrease of recall can be solved as the amount of training data increases (Table 2). We also retain the predicted binding score for cases where some tumor types do not have enough mutations to tolerate poor recall. Finally, the application of DeepHLApan to the mutations with pre-existing T-cell responses shows that it has a performance comparable to that of the state-of-the-art EDGE model in high-confidence neoantigen prediction under the expression threshold of TPM >2.

There are also some limitations of our study. First, DeepHLApan does not have a significantly improved performance on the published CD8⁺ T-cell epitopes compared with EDGE. One of the possible reasons for this result is

the limited number of immunogenic HLA-peptide pairs, which results in an immunogenicity model that is unable to classify all HLA-peptide pairs correctly. Another reason is the possibility that the datasets used for comparison are more suitable for EDGE. Another limitation of our study is that we consider all the HLA-peptide pairs as immunogenic if they have been validated to elicit T-cell activation at least once when training an immunogenic model. This assumption simplifies the evaluation of potential immunogenicity because whether the pHLA-matched TCR exists in the body is unknown. And with the alterations in the tumor microenvironment or immunoediting, the previous immunogenic neoantigens might be non-immunogenic. All these complex factors should be considered if we want to obtain accurate prediction. One of the solutions to this issue, and a direction for future research, would be to develop a model based on the pHLA-TCR pairs for more reliable immunogenicity prediction for specific individual. With that model, we could predict higher confidence neoantigens by taking advantage of whole-exon sequencing (call mutations) and RNA-Seq sequencing [evaluate gene expression and analyze the complementarity-determining regions 3 [CDR3] sequences by TRUST (55)] or TCR sequencing (call CDR3 sequences).

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.iedb.org/>, <http://www.ensembl.org>.

AUTHOR CONTRIBUTIONS

ZZ and SC conceived and supervised the study. JiaW and ZZ designed the tool. JinW, WW, and JZ wrote the scripts and trained the models. JinW, BZ, and WZ performed the data analysis. ZS, XG, and JiaW participated in designing and developing the algorithms. ZZ, JinW, and BZ drafted the manuscript. All authors have read and approved the final manuscript.

FUNDING

This work has been supported by the National Key R&D Program of China (Grant No. 2017YFC0908600), the National Natural Science Foundation of China (Grant No. 31971371), the Zhejiang Provincial Natural Science Foundation of China (Grant No. LY19H300003), and the Fundamental Research Funds for the Central Universities of China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

ACKNOWLEDGMENTS

We thank Prof. Jianning Song and Dr. Shutao Mei from Monash University for providing the independent validation

dataset. We also thank the Research and Service Center, College of Pharmaceutical Sciences, and the Real Doctor AI Research Center at Zhejiang University for providing computational support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2019.02559/full#supplementary-material>

Figure S1 | The distribution of binding data collected from the IEDB before and after data balance. **(a)** Before data balance, the distribution of 327,178 non-redundant HLA-peptide pairs, which covers 169 HLA alleles. The HLA alleles are displayed clockwise according to the HLA-peptide pairs they possess. The length of each HLA allele is proportionable to the number of HLA-peptide pairs. Red and green indicates positive and negative pairs, respectively. The histogram illustrates the length distribution of peptides (ranges from 8 to 15) of each HLA allele. **(b)** After data balance, the distribution of 437,077 HLA-pairs covering 81 HLA alleles. Only alleles with more than 3,000 HLA-peptide pairs are labeled.

Figure S2 | The HLA-peptide pair distribution of the removed alleles. These removed alleles are divided into four types: 62 had pure positive HLA-peptide pairs, one had pure negative pairs, 14 had positive pairs more than 5-fold of the negative pairs and four had negative pairs more than 5-fold of the positive pairs. The number in each cell represents the number of HLA-peptide pairs of each HLA allele.

Table S1 | The binding data collected from the IEDB. HLA-peptide without affinity are set to 0 in the "Affinity" column.

Table S2 | The detailed information of the removed alleles with their predicted binding scores and labels. In the "label" column, 0 stands for non-binding, and 1 stands for binding.

Table S3 | The binding data used for training the final binding model.

Table S4 | The immunogenicity data collected from the IEDB.

Table S5 | **(A)** HLA-peptide pairs created by the data from Tran et al., Gros et al., and Zacharakis et al. with their predicted binding scores and immunogenic scores. **(B)** HLA-peptide pairs created by the data from T Stronen et al. with their predicted binding scores and immunogenic scores.

Table S6 | Immunogenic HLA-peptide pairs derived from Gros et al., Tran et al., and Stronen et al. with their ranks predicted by DeepHLApan under TPM>2.

Table S7 | **(A)** The AUC performance of the binding model of DeepHLApan on the latest IEDB benchmark datasets compared with other tools. The best performance of each sub-dataset is highlighted in red. **(B)** The AUC performance of the binding model of DeepHLApan on the datasets collected from Mei et al. compared with other tools. The best performance of each sub-dataset is highlighted in red.

Table S8 | The validated neoantigens retrieved from Koşaloğlu-Yalçın et al. and the corresponding predicted immunogenic scores.

Table S9 | HLA-A02:01-restricted validated neoantigens retrieved from Koşaloğlu-Yalçın et al. and the immunogenic scores predicted by the HLA-A02:01-restricted model.

Table S10 | The predicted mutation score of four tools at TPM>0. Smaller the values of MHCflurry and NeMHCpan4 are and larger the values of EDGE and DeepHLApan are, stronger the binding probabilities.

Table S11 | The performance (AUC) of 4 tools on neoantigen prediction for each patient at different cut off of TPM. The best performance of each patient is highlighted in bold.

REFERENCES

- Lee C, Yelensky R, Jooss K, Chan TA. Update on tumor neoantigens and their utility: why it is good to be different. *Trends Immunol.* (2018) 39:536–48. doi: 10.1016/j.it.2018.04.005
- Lu YC, Robbins PF. Cancer immunotherapy targeting neoantigens. *Semin Immunol.* (2016) 28:22–7. doi: 10.1016/j.smim.2015.11.002
- Tran E, Robbins PF, Lu Y-C, Prickett TD, Gartner JJ, Jia L, et al. T-cell transfer therapy targeting mutant KRAS in cancer. *N Engl J Med.* (2016) 375:2255–62. doi: 10.1056/NEJMoa1609279
- Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature.* (2017) 547:222–6. doi: 10.1038/nature23003
- Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature.* (2017) 547:217–21. doi: 10.1038/nature22991
- Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med.* (2014) 371:2189–99. doi: 10.1056/NEJMoa1406498
- Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, et al. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science.* (2015) 348:124–8. doi: 10.1126/science.aaa1348
- Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with nivolumab. *Cell.* (2017) 171:934–49.e15. doi: 10.1016/j.cell.2017.09.028
- Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature.* (2014) 515:572–6. doi: 10.1038/nature14001
- Matsushita H, Vesely MD, Koboldt DC, Rickert CG, Uppaluri R, Magrini VJ, et al. Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoeediting. *Nature.* (2012) 482:400–4. doi: 10.1038/nature10755
- Zhou Z, Lyu X, Wu J, Yang X, Wu S, Zhou J, et al. TSNAD: an integrated software for cancer somatic mutation and tumour-specific neoantigen detection. *R Soc Open Sci.* (2017) 4:170050. doi: 10.1098/rsos.170050
- Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, et al. pVAC-Seq: a genome-guided *in silico* approach to identifying tumor neoantigens. *Genome Med.* (2016) 8:11. doi: 10.1186/s13073-016-0264-5
- Zhang J, Mardis ER, Maher CA. INTEGRATE-neo: a pipeline for personalized gene fusion neoantigen discovery. *Bioinformatics.* (2017) 33:555–7. doi: 10.1093/bioinformatics/btw674
- Rothbard JB, Taylor WR. A sequence pattern common to T cell epitopes. *EMBO J.* (1988) 7:93–100. doi: 10.1002/j.1460-2075.1988.tb02787.x
- Liu G, Li D, Li Z, Qiu S, Li W, Chao CC, et al. PSSMHCpan: a novel PSSM-based software for predicting class I peptide-HLA binding affinity. *Gigascience.* (2017) 6:1–11. doi: 10.1093/gigascience/gix017
- Luo H, Ye H, Ng HW, Sakkiah S, Mendrick DL, Hong H. sNebula, a network-based algorithm to predict binding between human leukocyte antigens and peptides. *Sci Rep.* (2016) 6:32115. doi: 10.1038/srep32115
- Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol.* (2017) 199:3360–8. doi: 10.4049/jimmunol.1700893
- Bui HH, Schiewe AJ, Von Grafenstein H, Haworth IS. Structural prediction of peptides binding to MHC class I molecules. *Proteins Struct Funct Genet.* (2006) 63:43–52. doi: 10.1002/prot.20870
- Mukherjee S, Bhattacharyya C, Chandra N. HLaffy: estimating peptide affinities for class-I HLA molecules by learning position-specific pair potentials. *Bioinformatics.* (2016) 32:2297–305. doi: 10.1093/bioinformatics/btw156
- Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol.* (2008) 4:e1000048. doi: 10.1371/journal.pcbi.1000048
- Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics.* (2012) 64:177–86. doi: 10.1007/s00251-011-0579-8
- Wu J, Zhao W, Zhou B, Su Z, Gu X, Zhou Z, et al. TSNADB: a database for tumor-specific neoantigens from immunogenomics data analysis. *Genom Proteom Bioinfo.* (2018) 16:276–82. doi: 10.1016/j.gpb.2018.06.003
- Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, et al. Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* (2017) 18:248–62. doi: 10.1016/j.celrep.2016.12.019
- Bassani-Sternberg M, Gfeller D. Unsupervised HLA peptidome deconvolution improves ligand prediction accuracy and predicts cooperative effects in peptide-HLA interactions. *J Immunol.* (2016) 197:2492–9. doi: 10.4049/jimmunol.1600808
- Kandalafi LE, Coukos G, Solleder M, Chong C, Bassani-Sternberg M, Pak HS, et al. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allosteric regulating HLA specificity. *PLOS Comput Biol.* (2017) 13:e1005725. doi: 10.1371/journal.pcbi.1005725
- Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity.* (2017) 46:315–26. doi: 10.1016/j.immuni.2017.02.007
- Bulik-Sullivan B, Busby J, Palmer CD, Davis MJ, Murphy T, Clark A, et al. Deep learning using tumor HLA peptide mass spectrometry datasets improves neoantigen identification. *Nat Biotechnol.* (2019) 37:55–63. doi: 10.1038/nbt.4313
- Ward JR, Gubin MM, Schreiber RD. The role of neoantigens in naturally occurring and therapeutically induced immune responses to cancer. *Adv Immunol.* (2016) 130:25–74. doi: 10.1016/bs.ai.2016.01.001
- Han Y, Kim D. Deep convolutional neural networks for pan-specific peptide-MHC class I binding prediction. *BMC Bioinformatics.* (2017) 18:585. doi: 10.1186/s12859-017-1997-x
- Vang YS, Xie X. HLA class I binding prediction via convolutional neural networks. *Bioinformatics.* (2017) 33:2658–65. doi: 10.1093/bioinformatics/btx264
- Hu Y, Wang Z, Hu H, Wan F, Chen L, Xiong Y, et al. ACME: pan-specific peptide-MHC class I binding prediction through attention-based deep neural networks. *Bioinformatics.* (2019). doi: 10.1093/bioinformatics/btz427. [Epub ahead of print].
- Liu Z, Cui Y, Xiong Z, Nasiri A, Zhang A, Hu J. DeepSeqPan, a novel deep convolutional neural network model for pan-specific class I HLA-peptide binding affinity prediction. *Sci Rep.* (2019) 9:794. doi: 10.1038/s41598-018-37214-1
- Balachandran VP, Luksza M, Zhao JN, Makarov V, Moral JA, Remark R, et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature.* (2017) 551:512–6. doi: 10.1038/nature24462
- Luksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyyov A, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature.* (2017) 551:517–20. doi: 10.1038/nature24473
- Kim S, Kim HS, Kim E, Lee MG, Shin E, Paik S, et al. Neopepsee: accurate genome-level prediction of neoantigens by harnessing sequence and amino acid immunogenicity information. Original article. *Ann Oncol.* (2018) 29:1030–6. doi: 10.1093/annonc/mdy022
- Jurtz VI, Jensen LE, Bentzen AK, Jespersen MC, Mahajan S, Vita R, et al. NetTCR: sequence-based prediction of TCR binding to peptide-MHC complexes using convolutional neural networks. *bioRxiv [Preprint].* (2018). doi: 10.1101/433706
- Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* (2015) 43:D405–12. doi: 10.1093/nar/gku938
- Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. *Nucleic Acids Res.* (2018) 46:D754–61. doi: 10.1093/nar/gkx1098
- Mei S, Li F, Leier A, Marquez-lago TT, Giam K, Croft NP, et al. A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief Bioinform.* (2019). doi: 10.1093/bib/bbz051. [Epub ahead of print].
- Koşaloğlu-Yalçın Z, Lanka M, Frentzen A, Logandha Ramamoorthy Premlal A, Sidney J, Vaughan K, et al. Predicting T cell recognition of MHC class I restricted neoepitopes. *Oncotarget.* (2018) 7:e1492508. doi: 10.1080/2162402X.2018.1492508

41. Tran E, Ahmadzadeh M, Lu YC, Gros A, Turcotte S, Robbins PF, Gartner JJ, Zheng Z, Li YF, Ray S, et al. Immunogenicity of somatic mutations in human gastrointestinal cancers. *Science*. (2015) 350:1387–90. doi: 10.1126/science.aad1253
42. Gros A, Parkhurst MR, Tran E, Pasetto A, Robbins PF, Ilyas S, et al. Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat Med*. (2016) 22:433–8. doi: 10.1038/nm.4051
43. Strønen E, Toebes M, Kelderman S, Buuren MM Van, Yang W, Rooij N Van, et al. Targeting of cancer neoantigens with donor-derived T cell receptor repertoires. *Science*. (2016) 352:1337–41. doi: 10.1126/science.aaf2288
44. Zacharakis N, Chinnasamy H, Black M, Xu H, Lu YC, Zheng Z, et al. Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer. *Nat Med*. (2018) 24:724–30. doi: 10.1038/s41591-018-0040-8
45. Cho K, van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv:1406:1078* (2014). doi: 10.3115/v1/D14-1179
46. Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate. *arXiv:1409:0473* (2014). Available online at: <https://arxiv.org/abs/1409.0473v1> (accessed September 01, 2014).
47. Raffel C, Ellis DPW. Feed-forward networks with attention can solve some long-term memory problems. *arXiv:1512:08756* (2016). Available online at: <https://arxiv.org/abs/1512.08756> (accessed December 29, 2015).
48. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*. (2009) 61:1–13. doi: 10.1007/s00251-008-0341-z
49. Crooks G, Hon G, Chandonia J, Brenner S. WebLogo: a sequence logo generator. *Genome Res*. (2004) 14:1188–90. doi: 10.1101/gr.849004
50. Liepe J, Marino F, Sidney J, Jeko A, Bunting DE, Sette A, et al. A large fraction of HLA class I ligands are proteasome-generated spliced peptides. *Science*. (2016) 354:605–10. doi: 10.1126/science.aaf4384
51. Zhang H, Lund O, Nielsen M. The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding. *Bioinformatics*. (2009) 25:1293–9. doi: 10.1093/bioinformatics/btp137
52. Rasmussen M, Fenoy E, Harndahl M, Kristensen AB, Nielsen IK, Nielsen M, et al. Pan-specific prediction of peptide–MHC class I complex stability, a correlate of T cell immunogenicity. *J Immunol*. (2016) 197:1517–24. doi: 10.4049/jimmunol.1600582
53. Donnell TJO, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J, et al. MHCflurry: open-source class I MHC binding tool MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst*. (2018) 7:129–32.e4. doi: 10.1016/j.cels.2018.05.014
54. Mylonas R, Beer I, Iseli C, Chong C, Pak H-S, Gfeller D, et al. Estimating the contribution of proteasomal spliced peptides to the HLA-I ligandome. *Mol Cell Proteom*. (2018) 17:2347–57. doi: 10.1074/mcp.RA118.000877
55. Li B, Li T, Pignon J-C, Wang B, Wang J, Shukla SA, et al. Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet*. (2016) 48:725–32. doi: 10.1038/ng.3581

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer SG and handling editor declared their shared affiliation.

Copyright © 2019 Wu, Wang, Zhang, Zhou, Zhao, Su, Gu, Wu, Zhou and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.