

RESEARCH ARTICLE

Open Access



MATHLA: a robust framework for HLA-peptide binding prediction integrating bidirectional LSTM and multiple head attention mechanism

Yilin Ye^{1,2†}, Jian Wang^{1†}, Yunwan Xu¹, Yi Wang¹, Youdong Pan¹, Qi Song¹, Xing Liu³ and Ji Wan^{1*} 

*Correspondence:

jyw@neocura.net

[†]Yilin Ye and Jian Wang have contributed equally to this work

¹ Shenzhen Neocura Biotechnology Co. Ltd., Shenzhen 518055, China
Full list of author information is available at the end of the article

Abstract

Background: Accurate prediction of binding between class I human leukocyte antigen (HLA) and neopeptide is critical for target identification within personalized T-cell based immunotherapy. Many recent prediction tools developed upon the deep learning algorithms and mass spectrometry data have indeed showed improvement on the average predicting power for class I HLA-peptide interaction. However, their prediction performances show great variability over individual HLA alleles and peptides with different lengths, which is particularly the case for HLA-C alleles due to the limited amount of experimental data. To meet the increasing demand for attaining the most accurate HLA-peptide binding prediction for individual patient in the real-world clinical studies, more advanced deep learning framework with higher prediction accuracy for HLA-C alleles and longer peptides is highly desirable.

Results: We present a pan-allele HLA-peptide binding prediction framework—MATHLA which integrates bi-directional long short-term memory network and multiple head attention mechanism. This model achieves better prediction accuracy in both fivefold cross-validation test and independent test dataset. In addition, this model is superior over existing tools regarding to the prediction accuracy for longer ligand ranging from 11 to 15 amino acids. Moreover, our model also shows a significant improvement for HLA-C-peptide-binding prediction. By investigating multiple-head attention weight scores, we depicted possible interaction patterns between three HLA I supergroups and their cognate peptides.

Conclusion: Our method demonstrates the necessity of further development of deep learning algorithm in improving and interpreting HLA-peptide binding prediction in parallel to increasing the amount of high-quality HLA ligandome data.

Keywords: Deep learning, HLA-peptide binding prediction, Cancer immunotherapy



Background

HLA-peptide binding is crucial for epitope presentation on the human cell surface and the elicitation of subsequent T cell immune response. In silico prediction of binary HLA-epitope binding or HLA-epitope binding affinity score has become one of the most essential criteria in target identification for a variety of applications in the immunotherapy [1, 2]. In general, the HLA-epitope prediction is largely dependent on state-of-the-art machine learning algorithms and substantial amount of data quantitating in vitro or in vivo HLA-epitope binding. In the past several years, a variety of deep learning models such as deep neural network [3], convolutional neural network [4] and recurrent neural network [5] have been developed for advancing HLA-peptide prediction over traditional machine learning algorithms [6]. Meanwhile, some studies weighted greatly on the data quality for improving prediction accuracy. For example, MHCflurry [3] constructed multiple models using the same architecture and discovered that the model derived from large-scale mass spectrometry data was able to outperform competing models. Recently, larger mass spectrometry data for 95 HLA class I alleles was generated, which further illustrated improving prediction accuracy by incorporating more data into the model [7]. However, data-dependent methods are confined by the limited number of alleles current technology can process and imbalance of data entries between three HLA class I supergroups (A, B, and C). As a result, data-dependent methods tend to result in lower prediction accuracy for HLA-C alleles due to smaller number of identified HLA-C ligands. Therefore, further refinement in deep learning architecture is required, which is especially the case for enhanced accuracy of pan-allele prediction tools. In contrast to the specialized model built by significant amount of training data of a fixed allele, pan-allele model is universal for predicting interaction between peptides and any HLA alleles. The principle of pan-allele model underlies that core sequences of HLA alleles can be explicitly outlined [8]. Thus, as HLA ligand sequence, the core HLA sequences can be encoded and fed into the learning algorithm for modeling HLA-peptide interaction. Compared to allele-specific methods, the pan-allele methods such as netMHCpan 4.0 [9] are more compelling to general users through higher compliance while retaining comparable performance [6]. However, prediction accuracy for a portion of HLA alleles, especially HLA-C alleles, are consistently far below average performance [6, 10]. In addition, since most natural ligands of HLAs are 8–11 amino acids in length [11], the limited amount of ligands of other lengths in the training dataset will affect the accuracy for predicting HLA-bound 12mer to 15mer [12]. Therefore, it is highly desirable to establish a more robust pan-allele model that is able to predict ligands of longer lengths and of cognate peptides of HLA-C alleles with higher accuracy.

Here we propose a novel deep learning HLA-epitope binding prediction method which takes advantage of the intrinsic ability of bi-directional LSTM to extract information from longer sequence and the ability of multiple head attention mechanism to capture contextual dependence from different angles. The proposed framework is more powerful in predicting the binding between HLA-C alleles and peptides. In addition, this framework is more robust than existing tools in predicting ligands ranging from 12 to 15 amino acids in length. Finally, this model can also help interpret the interaction between HLA alleles and peptides through multiple subspaces of interaction representation.

Methods

Datasets

Our training dataset is composed of data from IEDB [13], BD2013 dataset [14] and SysteMHC Atlas [15]. We only retained HLA class I ligands as well as peptides without post-translational modifications and ambiguous amino acids. In addition, ligands with relatively low confidence ($\text{prob} < 0.99$) in SysteMHC Atlas were excluded. All peptides in the final training dataset were between 8 and 15 amino acids in length. To balance the number between positive and negative datasets, we retrieved a companion decoy peptide from the host protein of the positive peptide from mass spectrometry data.

To convert qualitative affinity data to quantitative values, we applied a similar rule as MHCflurry [3]: positive-high, < 100 nM; positive, < 500 nM; positive-intermediate, < 1000 nM; positive-low, < 5000 nM; negative, > 5000 nM; MS-identified ligands, < 500 nM; decoys, > 5000 nM. In addition, we applied another set of rules to remove measurement redundancy for the same allele-peptide pair: keep the only data with “=” if other data were measured by inequality; if all the data are measured by “>”, the one with the greatest affinity values was retained; if all the data are measured by “<”, the one with lowest affinity value was then retained; and all the remaining data with contradictory measurements were discarded.

To facilitate model training, we normalized the original nanomolar affinity between 0 and 1.

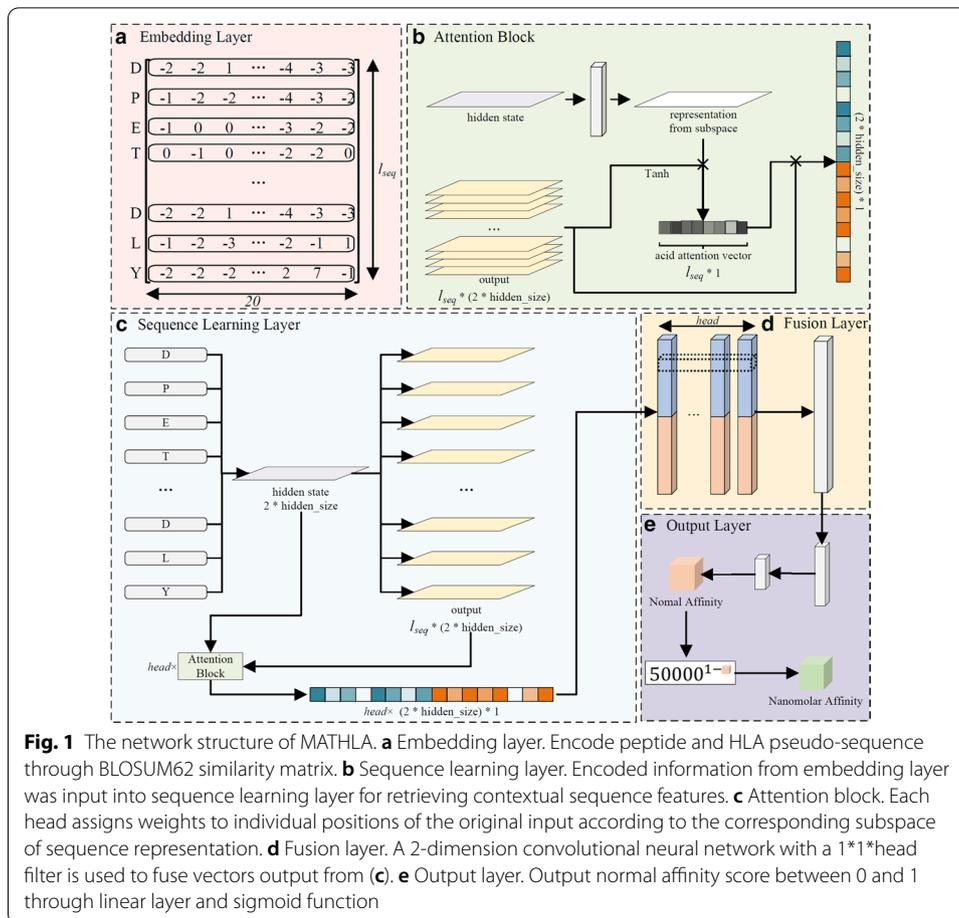
$$a_{normal} = 1 - \log_{50000}(a_{nM}) \quad (1)$$

where a_{normal} is normalized affinity and a_{nM} is the original nanomolar affinity value. The final training dataset is composed of 753,961 entries for 167 HLA class I alleles (53 HLA-As, 92 HLA-Bs and 22 HLA-Cs).

The positive data of the test dataset was compiled from a recent large-scale HLA class I ligandome data covering 95 HLA alleles. Data entries from 16 HLA alleles which were previously generated by the same group [16] and were included in the training dataset were first excluded. Next, we retained HLA-displayed ligands with length of 8 to 15 amino acids and removed those with post-translational modifications. To introduce negative data into the test dataset, we randomly sampled decoy peptide sequences, which were not included in the positive datasets from the host protein-coding transcripts of the positive peptides. For each positive peptide, 100 decoy sequences were generated correspondingly. Finally, after filtering out data entries overlapping with the training dataset, there are in total 140,232 positive peptides and 13,939,114 negative decoys in the test dataset.

Model structure

Each residue of peptide and HLA pseudo-sequence (retrieved from netMHCpan 4.0 [9]) is encoded to a similarity score vector according to the BLOSUM62 substitution matrix [17]. Different from many other methods with a predefined “padding” rule to ensure the equal dimension of input matrix during training, our model allows input sequences with flexible lengths. The encoded matrix with dimension $l_{seq} * 20$, where l_{seq} is the length of concatenated



sequence of peptide and HLA pseudo-sequence, is then input into sequence learning layer (Fig. 1a).

We chose the long short-term memory network [18] to model dependence between amino acid residues of peptides with flexible lengths. Compared to conventional recurrent neural network, LSTM network bearing gate control units (input gate, forget gate, and output gate) is able to learn dependency information between distant residues within peptide sequences more effectively.

To enhance the capability of our model to learn bidirectional dependence between n-terminal and c-terminal amino acid residues, bidirectional LSTM (bi-LSTM) was used [19]. By inputting both the forward and reverse sequences to LSTM networks with the same structure respectively, the outputs of LSTM h_t and h'_t at time t for forward and reverse sequence are derived and the HLA-peptide sequence at position t is represented as $hidden_t = [h_t, h'_t]$. Finally, the output of bi-LSTM is denoted as out^{lstm} (Fig. 1b).

To attend to peptide information at different positions from various subspaces of sequence representation, we applied multiple-head attention mechanism [20, 21] to the output of bidirectional LSTM.

$$W_i^{atten} = hidden^{lstm} \cdot W_i^{project} \quad (2)$$

$$Context_i = W_i^{atten} \cdot (\tan h(out^{lstm}))^T \tag{3}$$

$$Head_i = \frac{Context_i}{\sum_{k=0}^h Context_k} \cdot out^{lstm} \tag{4}$$

where $hidden^{lstm} \in R^{1 \times (hidden \times 2)}$ is a hidden state of the bi-LSTM network, $W_i^{project} \in R^{(hidden \times 2) \times (hidden \times 2)}$ are weights for projecting original hidden states to different representation subspaces. $W_j^{atten} \in R^{1 \times (hidden \times 2)}$ are attention weights, $(\cdot)^T$ represents transpose of a matrix. $Context_j \in R^{1 \times l_{seq}}$ is the context vector. $out^{lstm} \in R^{l_{seq} \times (hidden \times 2)}$ is the output of LSTM network and finally $Head_j \in R^{1 \times (hidden \times 2)}$ is the attention vector of the original sequence under the i th attention mechanism (Fig. 1c).

The concatenated output vector of forward and backward attention is combined. A 2-dimension convolutional neural network (2D CNN) with a head*1*1 filter is then applied to the combined vectors for a fusion vector by learning the weight of each head of attention (Fig. 1d).

$$Fusion = \tanh([Head_1, Head_2, \dots, Head_h] \cdot W_F) \tag{5}$$

where h stands for the number of head of multiple head attention module, $W_F \in R^{h \times 1 \times 1}$ is the filter of 2D CNN, and $Fusion \in R^{1 \times (hidden \times 2)}$ is the output vector after applying 2D CNN. Finally, a predicted value ranging from 0 to 1 is output by applying a linear layer with sigmoid activation function (Fig. 1e).

$$Output = \text{sigmoid}(Fusion \cdot W_o + b) \tag{6}$$

where $W_o \in R^{(hidden \times 2) \times 1}$ and b are the weight vector and bias for linear layer, respectively.

We randomly sampled 70% of the positive and negative data respectively from the training dataset to compile the training data. The remaining data were used as validation dataset for tuning hyperparameters.

To minimize the influence of outliers (noise) on model training, we employ an optimized Huber loss function [22] during training.

$$L(y, \bar{y}) = \begin{cases} 0.5 \times diff^2, & |diff| \leq \delta \\ \delta \cdot (|diff| - 0.5 \times \delta), & |diff| > \delta \end{cases} \tag{7}$$

$$diff = \begin{cases} \min(\bar{y} - y, 0), & \text{if measurement is } (<) \\ \max(\bar{y} - y, 0), & \text{if measurement is } (>) \\ y - \bar{y}, & \text{if measurement is } (=) \end{cases} \tag{8}$$

where \bar{y} and y are values of observed and predicted binding affinity respectively. When the inequality relationship between \bar{y} and y is not met, their difference ($diff$) will thus affect loss. Moreover, Huber loss will degenerate to MSE loss when $diff$ is less than expected value δ . Otherwise Huber loss uses linear errors to evaluate training losses, which is able to minimize the impact of hard-to-learn data on the performance of model training. RAdam is used for the optimization of model parameters. Compared to traditional Adam, RAdam [23] is able to adjust the variance of adaptive learning rate so as to

prevent model from converging to the local minimum. The batch size is chosen as 512 and the training stops as the loss over validation dataset shows no improvement after 5 consecutive epochs. The epoch number is set as 100. The learning rate is set as 0.001, and the dropout rate is set as 0.1.

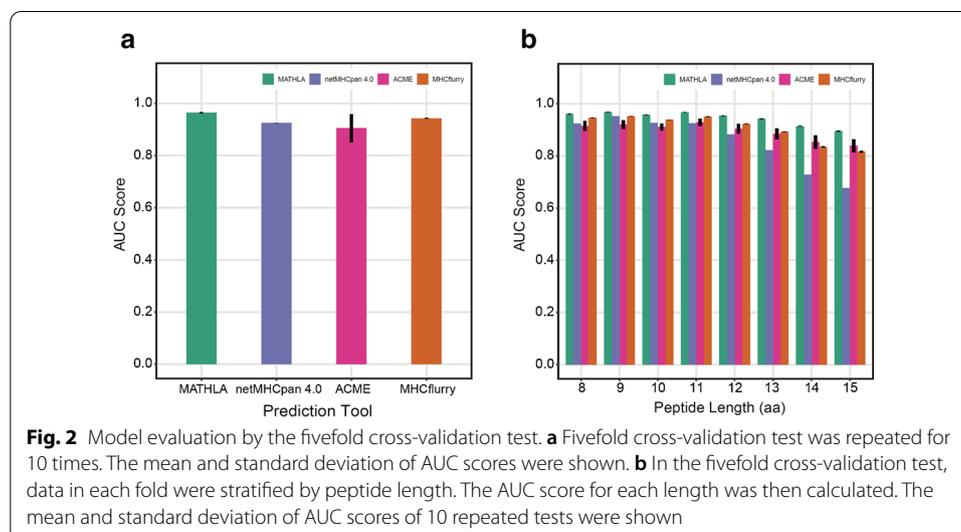
Results

Model evaluation based on fivefold cross-validation

To evaluate model performance and robustness, we conducted fivefold cross-validation test over training dataset. Area under the receiver operating characteristic curve (AUC) was used for model evaluation as well as model comparison against top-performed allele-specific and pan-allele models—MHCflurry, netMHCpan and ACME [24]. To ensure proper ratio of positive data to negative data in every fold of cross-validation, we separately divide positive and negative data into five individual subsets by random sampling. In each fold of cross-validation, four positive subsets and four negative subsets were pooled as training dataset whereas the remaining data were used as test dataset. The cross-validation tests were repeated for 10 times in order to calculate the mean and standard deviation. MATHLA achieves the best mean AUC score of 0.964, comparing to 0.945, 0.925 and 0.905 for netMHCpan 4.0, MHCflurry and ACME respectively (p values: $2.66e-16$, $9.06e-12$ and $4.91e-09$ respectively, one-sided t-test over AUC scores of 10 cross-validation repeats) (Fig. 2a).

The performance of MATHLA is more robust for longer HLA ligands

Most of previous methods tend to use sequence padding to handle flexible peptide lengths. Since LSTM is intrinsically designed for modelling longer sequence, we also examined the prediction performances of different tools over peptides ranging from 8 to 15 amino acids in length by fivefold cross-validation. The degrees of improvement by MATHLA over other models are positively correlated with the length of peptides (Fig. 2b). Especially for longer peptides of 12 to 15 amino acids in length, MATHLA achieves an average AUC score of 0.926 and shows 6.4%, 6.8% and 19.1% improvement



of average AUC score over ACME, MHCflurry and netMHCpan 4.0. In summary, the fivefold cross-validation results demonstrate that MATHLA outperforms state-of-the-art tools in both of model performance and robustness over variable lengths of ligands.

MATHLA outperforms existing pan-allele models on novel alleles

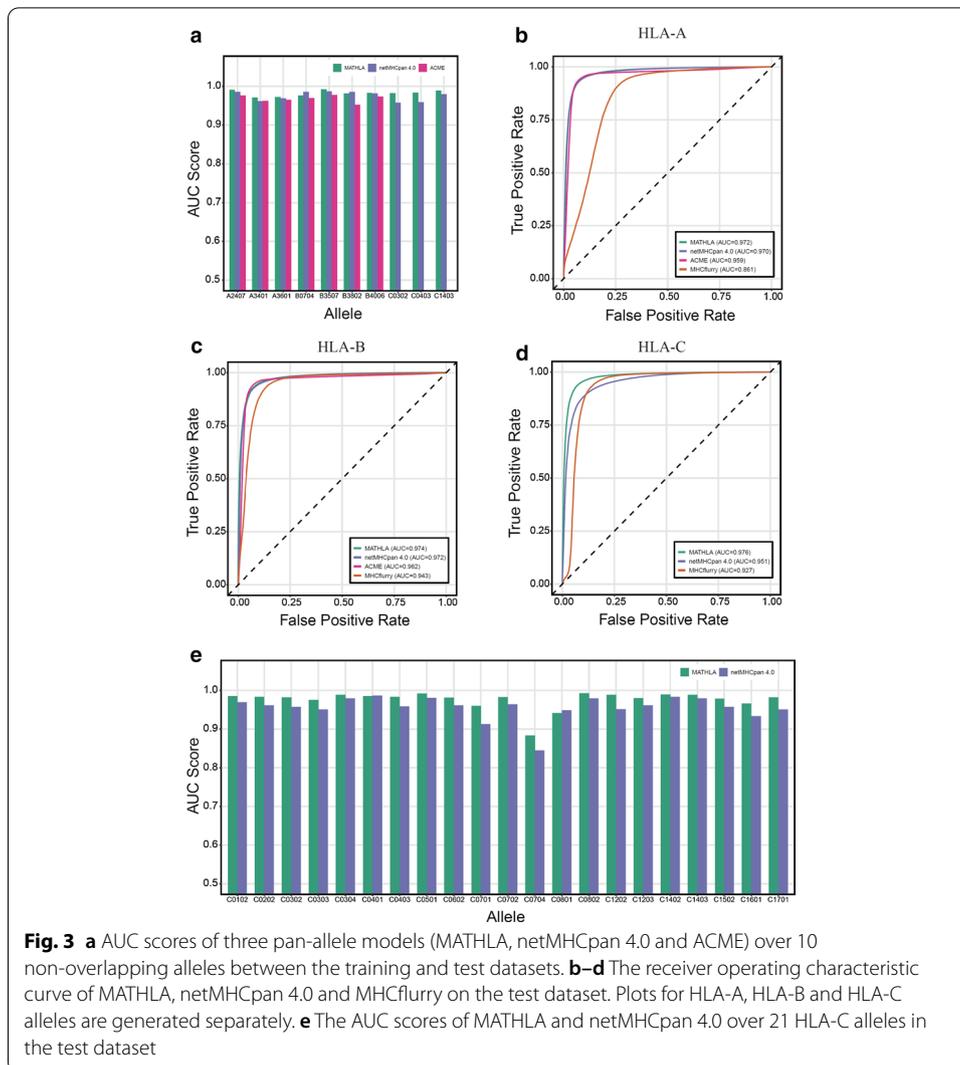
The foremost characteristic of a pan-allele model underlies its ability to accurately predict peptides bound to HLA alleles beyond the training dataset. To test the generalizability of MATHLA, we took the advantage of non-overlapping alleles between our training dataset and a set of mass spectrometry HLA ligandome data [7]. In total, 10 out of 95 alleles were used for assessing pan-allele model generalizability. Another two pan-allele models—netMHCpan 4.0 and ACME were used for model comparison regardless of whether these 10 alleles were included in their training datasets or not (only 7 HLA-A and HLA-B alleles are supported by ACME). In total, MATHLA outperforms netMHCpan 4.0 and ACME over 80% and 100% of non-overlapping alleles respectively. The average AUC of our model over 10 alleles reaches up to 0.982 which is higher than 0.975 of netMHCpan 4.0 (Fig. 3a). Relative to netMHCpan 4.0, it is noteworthy that performance advantage of MATHLA is more prominent for HLA-C alleles rather than HLA-A and HLA-B alleles. The average AUC score of three HLA-C alleles is 0.988 for MATHLA as compared to 0.965 for netMHCpan 4.0.

MATHLA improves accuracy over existing models for HLA-C alleles

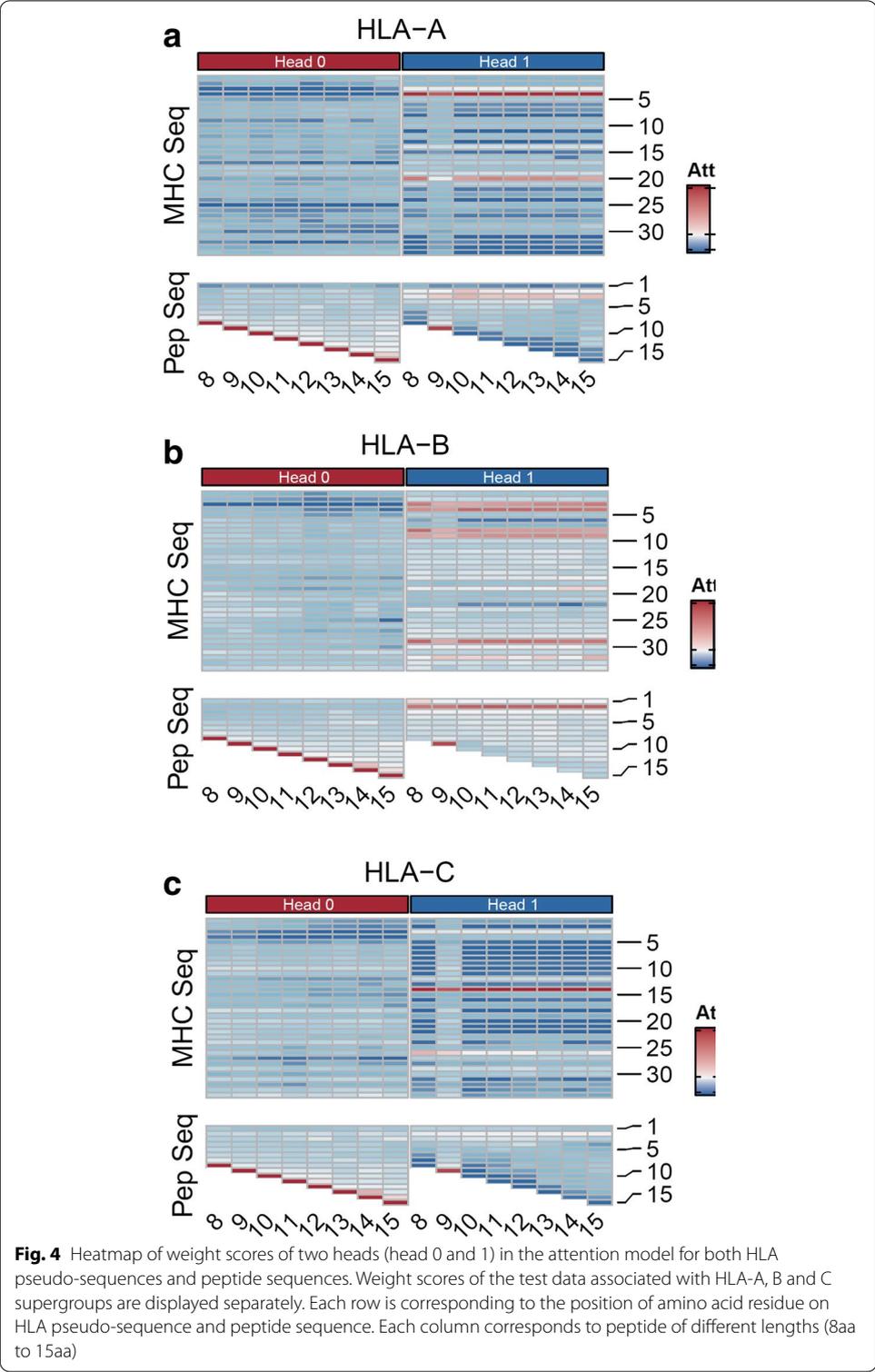
Inspired by the observation that MATHLA shows exceptional improvement for non-overlapping HLA-C alleles, we further compared the performance of our model over different supertypes of HLA I molecules. We separately calculated AUC scores of our model over all the HLA-A, B and C alleles in the test dataset. Pan-allele methods netMHCpan 4.0 and ACME, as well as allele-specific model MHCflurry, were used for model comparison. Although the AUC of MATHLA shows marginal enhancement over the top performed competing method netMHCpan for HLA-A and -B alleles, the AUC of MATHLA corresponding to HLA-C group is significantly improved over the competing models (0.976 for MATHLA, 0.951 for netMHCpan 4.0 and 0.927 for MHCflurry) (Fig. 3c). Moreover, we find that our model outcompetes netMHCpan 4.0 for 19 out of 21 (90.5%) individual HLA-C alleles. Since there are fewer data of HLA-C in the training dataset than those of HLA-A and HLA-B, we demonstrate that our model can outcompete both pan-allele and allele-specific models for HLA-C alleles in the context of limited number of training data.

MATHLA enables depiction of a variety of HLA-ligand binding patterns

To better understand the characteristics of model integrating bi-directional LSTM and multiple head attention mechanism, we investigated the attention weight scores corresponding to different supertypes of HLA class I molecules as well as ligands with different lengths. Previous motif analyses of HLA ligands revealed that the residue at the most C-terminal end was most likely to have recurring amino acids than other positions [25], which was confirmed by the consensus dominant weight score for the last residue of peptide sequence in the head 0 vectors across all three HLA class I supertypes (Fig. 4). On top of the consensus pattern in the head 0 vectors,



we observed more diversified patterns from the head 1 vectors. First, we found the weight distribution for peptide of 9 amino acids was distinct from peptides of all other lengths, in which the 9th position on the peptide is weighted dominantly. This pattern is consistent with previous finding that C-terminal residue is more important for the binding of 9-mer peptide than longer peptides [26]. Second, the attention weight scores of the second or third position on peptides bound by HLA-A and B are consistent with another known motif, while the corresponding positions for HLA-C ligands accounts for much less weights. This distinguished weight pattern of HLA-C ligands might explain why our model achieves greater advantages over other tools for HLA-C-peptide prediction. Collectively, our model demonstrates incorporating multiple head attention mechanism into LSTM network can capture HLA-supergroup-specific and peptide-length-specific information that enhances the robustness of MATHLA in HLA-ligand prediction.



Discussion

Machine learning based HLA-peptide binding prediction have been undergoing rapid

development in the past few years owing to the breakthrough of deep learning algorithm and emergence of large-scale mass spectrometry data. Meanwhile, prediction tools have been genuinely deployed in more and more clinical studies than before, due to the booming of cancer immunotherapy. The current criteria for selecting optimal tool in a real-world project mostly rely on the number of supporting alleles and average prediction accuracy measured by different metrics like AUC. However, in regard to prediction accuracy over individual alleles, there is no tool whose prediction accuracy is unambiguously higher over all the individual HLA alleles. This fact has posed several questions about HLA-peptide prediction on top of previous criteria like average prediction performance. First, for any given individual patient subject to target identification for cancer immunotherapy, how to select from a variety of existing tools the most accurate one for the given HLA alleles of this patient? This question becomes even more urgent and critical for patients carrying rare HLA alleles or alleles with limited amount of experimental data. Second, for longer ligands (11mer to 15mer) whose prediction accuracy is relatively lower, how to further improve the prediction accuracy given the fact the increasing mass spectrometry data alone can only provide limited prediction power.

Conclusions

Our model integrating bidirectional LSTM and multiple head attention mechanism has addressed these two questions by not only achieving prominent advantage in prediction accuracy for the HLA-C alleles but also attaining better prediction power for longer class I HLA ligands. Our work shows that advanced architecture of deep learning can provide an interpretive model to further improve and understand HLA-peptide binding prediction. We envision that introducing other alternative methods, such as self-attention mechanism and word2vec model, can provide better peptide representation to further improve prediction accuracy. Our framework will certainly benefit T cell-based vaccine development in treating cancers as well as for prevention of infectious diseases.

Abbreviations

HLA: Human leukocyte antigen; LSTM: Long short-term memory; AUC: Area under the receiver operating characteristic curve; CNN: Convolutional neural network; MSE: Mean squared error.

Acknowledgements

We are grateful to members of Neocura AI lab for helpful discussion and Junmei Hao for critical reading of the manuscript.

Authors' contributions

YL developed the model, implemented the software, and wrote the manuscript. JW organized the data and wrote the manuscript. YX, YW, YP, QS and XL revised the manuscript. JW designed the project and wrote the final manuscript. All authors have read and approved the manuscript.

Funding

None.

Availability of data and materials

The software is freely available at <https://github.com/MATHLAtools/>.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declared that they have no competing interests.

Author details

¹ Shenzhen Neocura Biotechnology Co. Ltd., Shenzhen 518055, China. ² School of Computer Science and Technology, Heilongjiang University, Harbin 150080, China. ³ The Center for Microbes, Development and Health, Key Laboratory of Molecular Virology and Immunology, Institut Pasteur of Shanghai, Chinese Academy of Sciences, Shanghai 200031, China.

Received: 30 July 2020 Accepted: 21 December 2020

Published online: 06 January 2021

References

- Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, Zhang W, Luoma A, Giobbie-Hurder A, Peter L, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature*. 2017;547(7662):217–21.
- Sahin U, Derhovanessian E, Miller M, Kloke BP, Simon P, Lower M, Bukur V, Tadmor AD, Luxemburger U, Schrors B, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature*. 2017;547(7662):222–6.
- O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. MHCflurry: open-source class I MHC binding affinity prediction. *Cell Syst*. 2018;7(1):129–132 e124.
- Vang YS, Xie X. HLA class I binding prediction via convolutional neural networks. *Bioinformatics*. 2017;33(17):2658–65.
- Chen B, Khodadoust MS, Olsson N, Wagar LE, Fast E, Liu CL, Muftuoglu Y, Sworder BJ, Diehn M, Levy R, et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat Biotechnol*. 2019;37(11):1332–43.
- Mei S, Li F, Leier A, Marquez-Lago TT, Giam K, Croft NP, Akutsu T, Smith AI, Li J, Rossjohn J, et al. A comprehensive review and performance evaluation of bioinformatics tools for HLA class I peptide-binding prediction. *Brief Bioinform*. 2019;21:1119–35.
- Sarkizova S, Klaeger S, Le PM, Li LW, Oliveira G, Keshishian H, Hartigan CR, Zhang WD, Braun DA, Ligon KL, et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat Biotechnol*. 2020;38(2):199.
- Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, Buus S, Nielsen M. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics*. 2009;61(1):1–13.
- 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(9):3360–8.
- Bonsack M, Hoppe S, Winter J, Tichy D, Zeller C, Kupper MD, Schitter EC, Blatnik R, Riemer AB. Performance evaluation of MHC class-I binding prediction tools based on an experimentally validated MHC-peptide binding data set. *Cancer Immunol Res*. 2019;7(5):719–36.
- Gfeller D, Guillaume P, Michaux J, Pak HS, Daniel RT, Racle J, Coukos G, Bassani-Sternberg M. The length distribution and multiple specificity of naturally presented HLA-I ligands. *J Immunol*. 2018;201(12):3705–16.
- 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*. 2016;8(1):33.
- Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res*. 2019;47(D1):D339–43.
- Kim Y, Sidney J, Buus S, Sette A, Nielsen M, Peters B. Dataset size and composition impact the reliability of performance benchmarks for peptide-MHC binding predictions. *BMC Bioinform*. 2014;15:241.
- Shao W, Pedrioli PGA, Wolski W, Scurtescu C, Schmid E, Vizcaino JA, Courcelles M, Schuster H, Kowalewski D, Marino F, et al. The SystemMHC Atlas project. *Nucleic Acids Res*. 2018;46(D1):D1237–47.
- Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, Stevens J, Lane W, Zhang GL, Eisenhaure TM, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. *Immunity*. 2017;46(2):315–26.
- Henikoff S, Henikoff JG. Amino-acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA*. 1992;89(22):10915–9.
- Gers FA, Schmidhuber J, Cummins F. Learning to forget: continual prediction with LSTM. *Neural Comput*. 2000;12(10):2451–71.
- Kaseliemi M, Doulamis N, Voulodimos A, Protopapadakis E, Doulamis A. Context aware energy disaggregation using adaptive bidirectional LSTM models. *IEEE Trans Smart Grid*. 2020;11:3054–67.
- Li J, Tu Z, Yang B, Lyu MR, Zhang T. Multi-head attention with disagreement regularization; 2018. *arXiv e-prints*. arXiv:1810.10183.
- Lin F, Zhang C, Liu S, Ma H. A hierarchical structured multi-head attention network for multi-turn response generation. *IEEE Access*. 2020;8:46802–10.
- Niu J, Chen J, Xu Y. Twin support vector regression with Huber loss. *J Intell Fuzzy Syst*. 2017;32(6):4247–58.
- Liu L, Jiang H, He P, Chen W, Liu X, Gao J, Han J. On the variance of the adaptive learning rate and beyond; 2019. *arXiv e-prints*. arXiv:1908.03265.
- Hu Y, Wang Z, Hu H, Wan F, Chen L, Xiong Y, Wang X, Zhao D, Huang W, Zeng J. ACME: pan-specific peptide-MHC class I binding prediction through attention-based deep neural networks. *Bioinformatics*. 2019;35(23):4946–54.
- Bassani-Sternberg M, Chong C, Guillaume P, Solleder M, Pak H, Gannon PO, Kandalaf LE, Coukos G, Gfeller D. Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allosteric regulating HLA specificity. *PLoS Comput Biol*. 2017;13(8):e1005725.
- Chen Y, Sidney J, Southwood S, Cox AL, Sakaguchi K, Henderson RA, Appella E, Hunt DF, Sette A, Engelhard VH. Naturally processed peptides longer than nine amino acid residues bind to the class I MHC molecule HLA-A2.1 with high affinity and in different conformations. *J Immunol*. 1994;152(6):2874–81.

Publisher's Note

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

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

