Briefings in Bioinformatics, 2025, **26(3)**, bbaf246
https://doi.org/10.1093/bib/bbaf246
Problem Solving Protocol

LightCTL: lightweight contrastive TCR-pMHC specificity learning with context-aware prompt

Fei Ye, Mao Chen, Yixuan Huang, Ruihao Zhang, Xuqi Li, Xiuyuan Wang, Sanyang Han, Lan Ma, Xiao Liu*

Institute of Biopharmaceutical and Health Engineering, Tsinghua ShenZhen International Graduate School, Tsinghua University, Lishui Road, Nanshan District, Shenzhen, Guangdong Province 518055, China

*Corresponding author. Tsinghua ShenZhen International Graduate School, Tsinghua University, Shenzhen 518055, China. E-mail: liuxiao@sz.tsinghua.edu.cn

Abstract

Identification of T cell receptor (TCR) specificities for antigens from large-scale single-cell or bulk TCR repertoire data plays a vital role in disease diagnosis and immunotherapy. *In silico* prediction models have emerged in recent years. However, the generalizability and transferability of current computational models remain significant hurdles in accurately predicting TCR–pMHC binding specificity, primarily due to the limited availability of experimental data and the vast diversity of TCR sequences. In this paper, we propose a lightweight contrastive TCR–pMHC learning with context-aware prompts, named LightCTL, to infer TCR–pMHC binding specificity. For each TCR and peptide-MHC sequence, we utilize a TCR encoding module and a pMHC encoding module to transform them into latent representations. Specifically, we introduce a contrastive TCR–pMHC learning paradigm to enhance the generalization ability of TCR–pMHC binding specificity prediction by learning the matching relationship between TCR–pMHC and MHC-peptide. We fuse the TCR and pMHC latent representations and employ a novel context-aware prompt module to consider the varying importance of different feature maps. Compared with existing methods, LightCTL substantially improves the accuracy of predicting TCR–pMHC binding specificity. Moreover, comparative experiments across eight independent datasets demonstrate the generalization ability of LightCTL, showing superior performance for predicting unknown TCR–pMHC pairs. Finally, we assess LightCTL's efficacy across different TCR sequence lengths and distinct unseen epitopes, as well as estimate cytomegalovirus-specific TCR diversity and clone frequency from peripheral TCR repertoire data. Overall, our findings highlight LightCTL as a versatile analytical method for advancing novel T-cell therapies and identifying novel biomarkers for disease diagnosis.

Keywords: lightweight; contrastive learning; context-aware prompt; TCR–pMHC binding specificity

Introduction

T cell receptors (TCRs) recognize and bind to specific antigenic peptides presented by the major histocompatibility complex (MHC) on the surface of antigen-presenting cells to form the TCR–pMHC complex, which induces T-cell activation and adaptive immune effector function [1, 2] (Fig. 1a). The TCR is a heterodimer consisting of α chain and β chain, and it exhibits a wide sequence diversity, with estimates ranging from 10^{15} to 10^{61} distinct TCR sequences that may be generated [3, 4]. This high diversity enables T cells to recognize an abundance of epitopes presented by different MHC alleles [5]. This diversity arises primarily from the plasticity of the three complementarity-determining region (CDR) loops—CDR1, CDR2, and CDR3—on the α and β chains, where CDR1 and CDR2 mainly mediate contact with MHC and CDR3 is responsible for binding to antigenic peptides [6]. Thus, the CDR3 sequence is the most diverse and a major determinant of antigen-binding specificity. Studying TCR-antigen binding specificity is of great significance in tumor immunotherapy, autoimmune antigen discovery, and vaccine design. Numerous experimental techniques have been developed to determine TCR–pMHC binding specificity, such as tetramer analysis [7], tetramer-associated TCR sequencing [8], and T-scan [9]. However, these traditional biological experimental methods are time-consuming, costly, and inefficient for validation. Therefore, computational approaches play a vital role in efficiently discovering specific TCRs associated with antigens.

The collection and construction of large-scale experimentally validated TCR–pMHC databases, such as VDJDB [10], McPAS-TCR [11], IEDB [12], and PIRD [13], have significantly advanced the development of *in silico* methods for predicting TCR–pMHC binding specificity. GLIPH [14] was the first computational method to predict TCR sequences that recognize pathogens or tumor *de novo* antigenic peptides based on the idea that TCRs with similar CDR3 sequences are likely to target antigenic epitopes with similar sequences, laying the groundwork for computationally based methods aimed at predicting TCR specificity. Subsequently, a variety of computational methods for predicting TCR–pMHC specificity have emerged (i.e. Gaussian Process [15, 16], Random Forests [17], Bayesian Probability Model [18], Energy Model [19]). Nevertheless, these approaches are only effective with small datasets

Received: August 29, 2024. **Revised:** March 15, 2025. **Accepted:** May 2, 2025

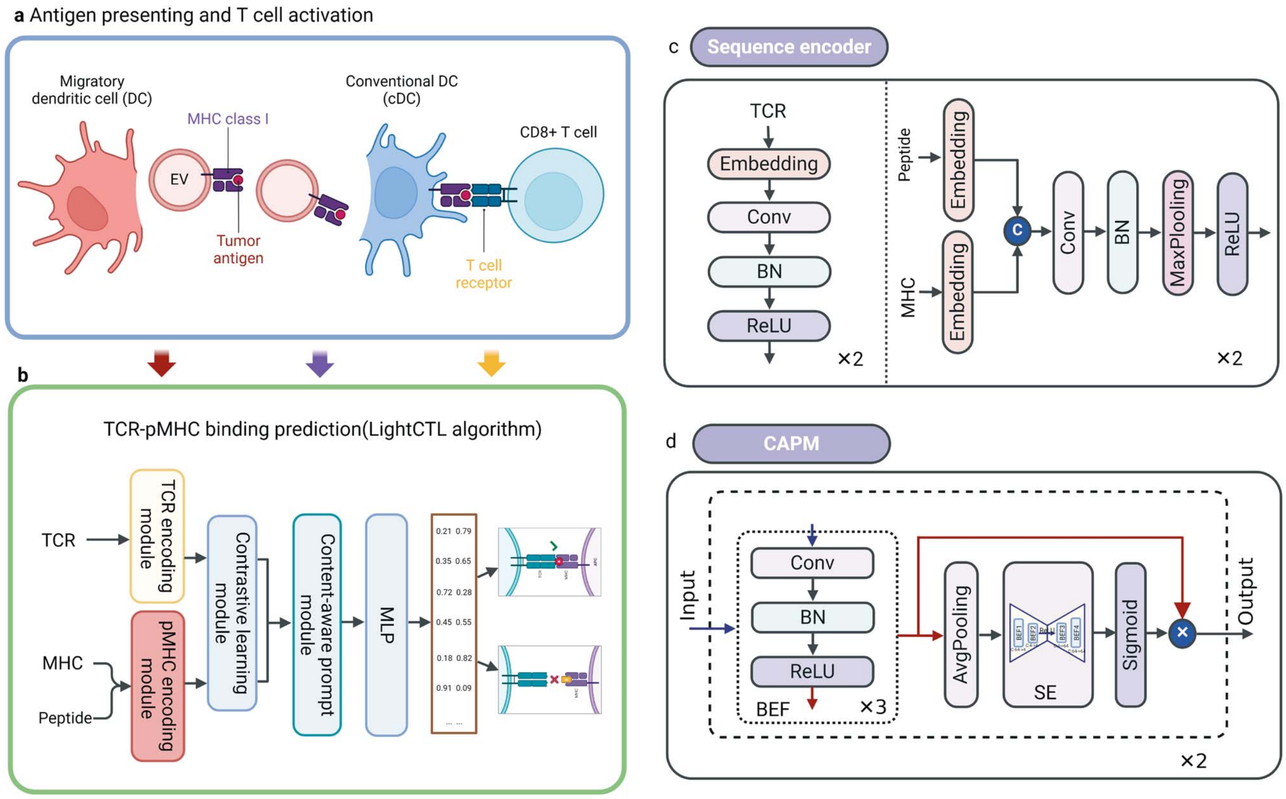


Figure 1. The architectures of LightCTL. (a) Antigen presentation and T cell activation. (b) The flowchart of LightCTL. (c) Details of the TCR and pMHC encoding module. (d) Details of the CAPM. Created with BioRender.com

and are incapable of predicting TCR sequences not included in the training set, which severely limits their practical applicability.

Deep learning has further promoted the development of TCR–pMHC binding specificity prediction and acquired encouraging performance due to its superior feature learning ability. These methods can be broadly categorized into structure-based methods and sequence-based methods. Structure-based methods are mainly based on Alphafold or its variants to predict the structure of the full-length TCR sequence, which is then used to predict TCR antigen specificity. Examples of such methods include Modeller and ColabFold [20], AlphaFold [21], TCRmodel [22], TCRmodel2 [23], and DeepAIR [24]. Additionally, there are customized algorithms based on TCR structure information. For example, TEIM-Res [25], which uses transfer learning, has been suggested to characterize the interaction conformation between TCRs and epitopes at the residue level. However, predicting antigen-specific TCRs based on their three-dimensional structure is limited by several factors, including the scarcity of experimentally determined TCR–pMHC complex structures, the high demand for hardware resources, and the large number of TCRs and antigenic targets in each individual. Consequently, sequence-driven strategies for predicting TCR antigen specificity have become a prevalent method in research. Over the past 5 years, numerous methods based on convolutional neural networks (CNNs) and recurrent neural networks have been applied to this prediction task, such as NetTCR [26], pMTnet [27], TITAN [28], DLpTCR [29], AttnTAP [30], etc. These methods demonstrate the validity of computational prediction of TCR antigen binding specificity. Recently, transformer-based methods and meta-learning approaches have been introduced for TCR–pMHC binding specificity prediction, such as MITNet [31], TPBTE [32], ProtBer [33], SC-AIR-BERT [34], ATMTCR [35], Pan-pep [36], Multimodal-AIR-BERT [37], etc. However, these methods often fail to fully consider the binding patterns among TCR, peptides, and MHC. Furthermore, the existing mainstream transformer-based methods learn the importance of different amino acids in the sequence using the self-attention mechanism, but do not take into account the degree of focus of various amino acid clusters in the sequence. Therefore, they do not migrate well to new datasets, and there is an excellent challenge in efficiently predicting the binding specificity between TCR and pMHC.

In this work, we propose LightCTL, a lightweight contrastive learning framework for predicting TCR–pMHC binding specificity, which leverages the CDR3 region of the TCR β chain (referred to as TCR in the following text), as well as peptide and MHC class I information. We systematically validated LightCTL using a large amount of independent validation data and achieved superior predictive performance compared with previous methods. To demonstrate the application of LightCTL, we applied it to TCR sequencing data from human peripheral blood samples, both with and without cytomegalovirus (CMV) infection. We analyzed the diversity, clonal frequency, and other characteristics of CMV-specific TCR sequences. LightCTL effectively addresses the long-standing challenge of TCR–pMHC pair prediction and could serve as a foundation for developing biomarkers for disease diagnosis and predicting responses to immunotherapy.

Results

Overview of LightCTL

We employed a novel lightweight contrastive TCR–pMHC learning with a context-aware prompt, LightCTL, to predict the TCR–pMHC binding specificity. The framework of LightCTL is shown in Fig. 1b. First, we converted the TCR, peptide, and MHC into computer-readable numerical information, which serves as the input to our model. Concretely, the Atchley factor is used for the TCR sequence, BLOSUM50 for the peptide sequence, and the MHC pseudo sequence. In particular, the MHC pseudo-sequence with 34 amino acids is generated by netMHCpan [38]. Next, a TCR encoding module and a pMHC encoding module are designed to extract learnable numeric embedding of TCR, peptide, and MHC and their shallow features, respectively (Fig. 1c). Unlike using a transformer as a feature extractor in self-supervised comparative learning, we designed a CNN-based extractor to extract shallow information of TCR, peptide, and MHC, which is able to reduce the number of parameters. Subsequently, a contrastive TCR–pMHC learning paradigm is proposed to learn the matching binding pattern of TCR–pMHC and peptide-MHC. Differing from prior self-supervised contrastive learning methods, LightCTL can learn implicit feature representation by learning the matching relationships between TCRs and pMHCs. Therefore, it pays more attention to solving the generalization problem by considering the diversity of sequences. Furthermore, the embedded TCR and pMHC features are fused and then input into a context-aware prompt module (CAPM) to extract more global feature information (Fig. 1d). The CAPM considers the importance of different amino acid groups in the TCR and pMHC sequences for TCR–pMHC binding specificity prediction. Finally, a multilayer perceptron (MLP) consisting of two fully connected layers is used to output the predictions of TCR antigen specificity. More details of the pipeline are provided in the Methods section.

Comparison of different training and encoding methods

Data were collected from two databases, including McPAS-TCR [11] and YFV [39] (The detailed procedures for constructing the training and validation datasets are described in the Supplementary Method 1.1). We train our model on the McPAS-TCR dataset and verify its performance on the YFV dataset, selecting the parameters with the best results on YFV data (external validation). The reason is that we want to see how well the algorithm learns when the data features of the two datasets are distributed differently (herein, McPAS-TCR and YFV datasets). To evaluate the performance of our training strategy, we compare it with the other two training methods on the McPAS-TCR dataset, which is 10-fold cross-validation and repeated hold-out with 80% in the training and the remaining 20% in the validation set. In both 10-fold cross-validation and repeated hold-out experiments, the model shows relatively stable performance with the average AUC of 0.9356 and 0.9371, respectively (Fig. 2a and 2b). In contrast, LightCTL with external validation outperformed them with an AUC of 0.9580 (Fig. 2c). Repeated hold-out experiment using the reservation method effectively considered the randomness in data partitioning, and the 10-fold cross-validation not only accounted for data partitioning randomness but also captured the feature distribution of the entire dataset more accurately. However, these approaches cannot be directly transferred to new datasets due to probable variations in feature distributions among datasets, which is an essential factor under our consideration. Before model training, TCR, peptide, and MHC class I pair sequences are required to be converted into discrete numerical vectors; therefore, selecting the most appropriate word2vector method is significant to construct a highly accurate model. Five encoding methods (see Supplementary Method 2.1) were conducted on the YFV dataset, including one-hot, BLOSUM50, Chem, Atchley factor, and reduced dimensionality amino acid indexes (AA_index_PCA). We can see that the Atchley factor can achieve the best overall performance with AUC up to 0.96 (Fig. 2d).

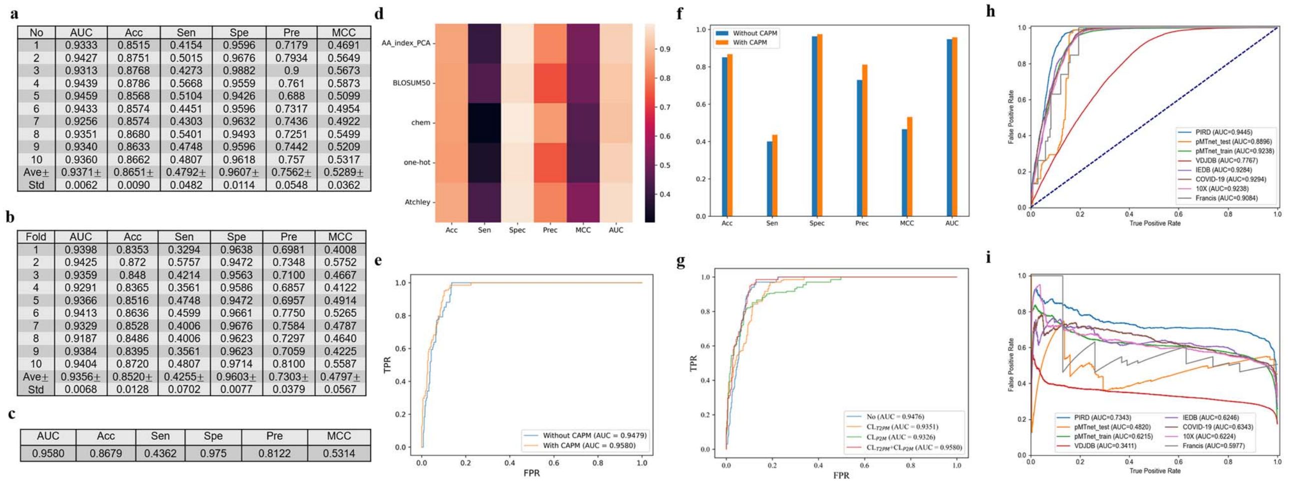
The influence of the CAPM and the contrastive TCR–pMHC learning

The CAPM makes the model pay more attention to the discriminative amino acid group features. We aim to provide those features with different weights according to the degree of attention. To verify the effectiveness of the CAPM in the model, we compare the LightCTL to the CNN without the CAPM. To ensure a fair comparison, both approaches use identical training and testing datasets, and all network settings are consistent. From Fig. 2e, we can see that the AUC value of LightCTL with a CAPM increases by 1.01%. In addition, all metrics of LightCTL are superior to the model without a CAPM (Fig. 2f). In detail, Acc, Sen, Spe, Pre, and MCC of LightCTL are improved from 0.8514 to 0.8679, 0.3988 to 0.4362, 0.9632 to 0.975, 0.7283 to 0.8122, and 0.464 to 0.5314, respectively. The results show that the CAPM can effectively help capture TCR–pMHC binding specificity.

To simulate the biology process of antigen processing and T cell recognition, we proposed a contrastive TCR–pMHC learning paradigm to learn the pattern of pMHC recognized by TCR and peptide binding to MHC. To evaluate the ability of the contrastive TCR–pMHC learning paradigm to learn generalization features, we compare the experimental results of the two contrastive loss functions (CL_{T2PM} and CL_{P2M}) in different combinations. The AUC value is improved from 0.9326 to 0.9580 when both contrastive loss functions are used, which demonstrates that CL_{T2PM} and CL_{P2M} can learn implicit binding features between TCRs and pMHCs, as well as between peptides and MHCs (Fig. 2g). A possible reason is that the binding feature between peptide and MHC provided by CL_{P2M} drives the learning of the binding pattern between TCR–pMHC based on CL_{T2PM}.

Evaluating the predictive performance of LightCTL on unseen TCR–pMHC pairs

The primary goal of LightCTL is to address the challenge of algorithm generalization. To demonstrate that LightCTL can not only transfer to new data sets that are smaller than the training data set but also perform better on larger, unseen datasets, we evaluate its generalization ability on eight independent datasets: PIRD [13], VDJDB [10], pMTnet_train [27], pMTnet_test [27], IEDB [12], 10X [40], COVID-19 [41], and Francis's data [42]. After data preprocessing, these eight datasets are all unseen TCR–pMHC pairs, which do not overlap with the training set (detailed procedures for constructing the testing datasets are described in Supplementary Method 1.2). Among them, VDJDB and pMTnet_train have more samples compared with McPAS-TCR and YFV. In contrast, pMTnet_test, IEDB, and Francis's data are smaller than McPAS-TCR, while PIRD, COVID-19, 10X, and McPAS-TCR have comparable sample sizes (Supplementary Table 1). Specifically, we choose the best models trained on the validation dataset to be tested directly on the eight external datasets. Our model achieves superior AUC results of 0.9445, 0.9238, 0.8896, 0.9284, 0.9238, 0.9084, and 0.9294 on PIRD, pMTnet_train, pMTnet_test, IEDB, 10X, Francis's data, and COVID-19, respectively. However, LightCTL performs slightly worse on the VDJDB dataset (Fig. 2h), and it also needs to make more efforts with PR metrics (Fig. 2i). To assess the stability and reliability of LightCTL, we extracted the model evaluation metrics and their confidence intervals, visualizing the performance through error bars and dynamic circles. We observe that LightCTL demonstrates strong performance on three key metrics (AUC, Acc, and AUPR). Additionally, LightCTL also



a

No	AUC	Acc	Sen	Spe	Pre	MCC
1	0.9333	0.8515	0.4154	0.9596	0.7179	0.4691
2	0.9427	0.8751	0.5015	0.9676	0.7934	0.5649
3	0.9313	0.8768	0.4273	0.9882	0.9	0.5673
4	0.9439	0.8786	0.5668	0.9559	0.761	0.5873
5	0.9459	0.8568	0.5104	0.9426	0.688	0.5099
6	0.9433	0.8574	0.4451	0.9596	0.7317	0.4954
7	0.9256	0.8574	0.4303	0.9632	0.7436	0.4922
8	0.9351	0.8680	0.5401	0.9493	0.7251	0.5499
9	0.9340	0.8633	0.4748	0.9596	0.7442	0.5209
10	0.9360	0.8662	0.4807	0.9618	0.757	0.5317
Ave± Std	0.9371± 0.0062	0.8651± 0.0090	0.4792± 0.0482	0.9607± 0.0114	0.7562± 0.0548	0.5289± 0.0362

b

Fold	AUC	Acc	Sen	Spe	Pre	MCC
1	0.9398	0.8353	0.3294	0.9638	0.6981	0.4008
2	0.9425	0.872	0.5757	0.9472	0.7348	0.5752
3	0.9359	0.848	0.4214	0.9563	0.7100	0.4667
4	0.9291	0.8365	0.3561	0.9586	0.6857	0.4122
5	0.9366	0.8516	0.4748	0.9472	0.6957	0.4914
6	0.9413	0.8636	0.4599	0.9661	0.7750	0.5265
7	0.9329	0.8528	0.4006	0.9676	0.7584	0.4787
8	0.9187	0.8486	0.4006	0.9623	0.7297	0.4640
9	0.9384	0.8395	0.3561	0.9623	0.7059	0.4225
10	0.9404	0.8720	0.4807	0.9714	0.8100	0.5587
Ave± Std	0.9356± 0.0068	0.8520± 0.0128	0.4255± 0.0702	0.9603± 0.0077	0.7303± 0.0379	0.4797± 0.0567

c

AUC	Acc	Sen	Spe	Pre	MCC
0.9580	0.8679	0.4362	0.975	0.8122	0.5314

Figure 2. Evaluation of LightCTL in different settings. (a) Results of hold-out running 10 times. (b) Results of 10-fold cross-validation. (c) Results of external validation. (d) Heatmap of LightCTL with different encoding methods for TCR sequences. (e) ROC curves of LightCTL with and without the CAPM. (f) Comparison of LightCTL with and without CAPM. (g) ROC curves of LightCTL with different combinations of contrastive TCR–pMHC learning paradigm. (h) and (i) ROC and PR curves of LightCTL on eight independent datasets, respectively.

performs well on the PIRD, pMTnet_train, IEDB, COVID-19, and 10X datasets (Supplementary Figure 2).

LightCTL achieves state-of-the-art performance to predict TCR–pMHC binding specificity

To further validate the generalization of LightCTL and its predictive ability for unknown TCR-antigen pairs by comparison with state-of-the-art (SOTA) deep learning methods, we compare LightCTL with several representative deep learning frameworks (BiLSTM [43], CNN [44], Transformer [45]) and published TCR–pMHC prediction methods (pMTnet [27], ATMTCR [35], PISTE [46], TABR-BERT [47], EPACT [48]). Compared with these methods, LightCTL showed significant improvements: 2.72%–19.51% on PIRD, 3.08%–36.24% on pMTnet_train, 5.1%–33.52% on pMTnet_test, 6.2%–41.39% on IEDB, 6.78%–40.01% on 10X, and 7.7%–54.42% on COVID-19. LightCTL slightly underperforms ATMTCR on the VDJDB dataset and performs slightly worse than Transformer on Francis's data (Fig. 3). In addition, we calculate the PR value and discover a similar conclusion as the AUC metric with imbalanced positive and negative sample sizes, which LightCTL ranks second on the VDJDB and Francis's datasets and performed best on the other datasets (Supplementary Figure 3a). Besides, the values of ACC display that LightCTL outperforms other methods on most datasets (Supplementary Figure 3b). Overall, these comparisons have shown that our model has better generalization capabilities and achieves SOTA performance in TCR–pMHC binding specificity prediction. Further, the computational cost between LightCTL and existing methods is compared, and it is found that LightCTL has fewer parameters than PISTE and TABR-BERT, and its FLOPS are lower than those of Transformer and TABR-BERT. Additionally, the inference time of LightCTL is faster than that of Transformer, PISTE, TABR-BERT, and EPACT (Supplementary Table 2). While LightCTL may not be the most computationally efficient model, it is more lightweight than most transformer-based approaches. Furthermore, LightCTL has potential applications in the discovery of key contact sites in TCR–pMHC complexes. We collected 17 TCR–pMHC complexes from the STCRDab databases [49] and calculated the effect of different TCR–epitope residue pairs on the expected results to explore their relationship with the contact distance between residues on the TCR–pMHC structure. We identified contributing residue pairs in the presence of direct contact residue pairs in all TCR–pMHC complexes. We also observed in most of the TCR–pMHC complexes that the residue pairs with the highest contribution values were direct contacts, suggesting these could be potential key sites for TCR–epitope binding (Supplementary Figure 4).

Effect of the TCR sequence length and unique epitopes on LightCTL

To further evaluate the predictive ability of LightCTL for different TCR sequence lengths, we compare it with advanced deep learning methods using YFV data across varying sequence lengths. The TCR sequence length in the YFV dataset ranges from 10 to 22 amino acids (aa). Overall, LightCTL performs quite promising for the middle TCR sequence lengths between 13aa and 18aa, while it is not as good for lengths >19aa (Fig. 4a). We infer that LightCTL might be more sensitive to sample sizes, as the number of TCR–pMHC pairs in the two intervals [10-12] aa and [19-22] aa is significantly lower than in the other two intervals.

To evaluate the performance of LightCTL for different epitope peptides, we separately calculated its prediction results for the two peptides in the YFV dataset (NLVPMVATV and LLWNGPMAV). As shown in Fig. 4b, LightCTL achieved good results for both peptides, with an AUC of 0.9432 for NLVPMVATV and 0.7291 for LLWNGPMAV. Interestingly, LightCTL's prediction for the two peptides varies greatly, which may be related to the sample size of the peptides ($n = 1506$ for NLVPMVATV and $n = 516$ for LLWNGPMAV). We also evaluated the performance of LightCTL on unseen peptides (top 10 or all) of four independent datasets (i.e., 10X, COVID-19, IEDB, and Francis's data). Our results show that LightCTL demonstrates strong predictive performance on most unknown peptides (Fig. 4c).

LightCTL exhibits high performance in virus immune recognition

We further investigated the applications of LightCTL in virus immune recognition for the CMV study. We collected the CMV viral TCR bulk sequencing data published by Emerson *et al.* [50]. The data of 666 samples with HLA information were processed in our demonstration. After the removal of subjects

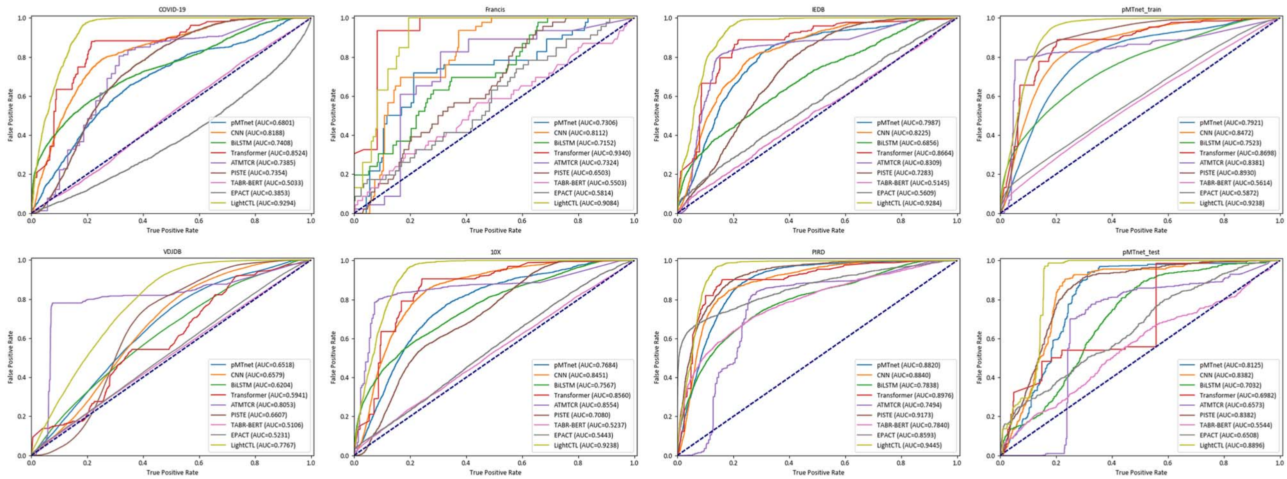


Figure 3. ROC curves of LightCTL for TCR–pMHC binding specificity prediction on eight independent datasets (PIRD, VDJDB, pMTnet_train, pMTnet_test, IEDB, 10X, COVID-19, and Francis’s data), respectively.

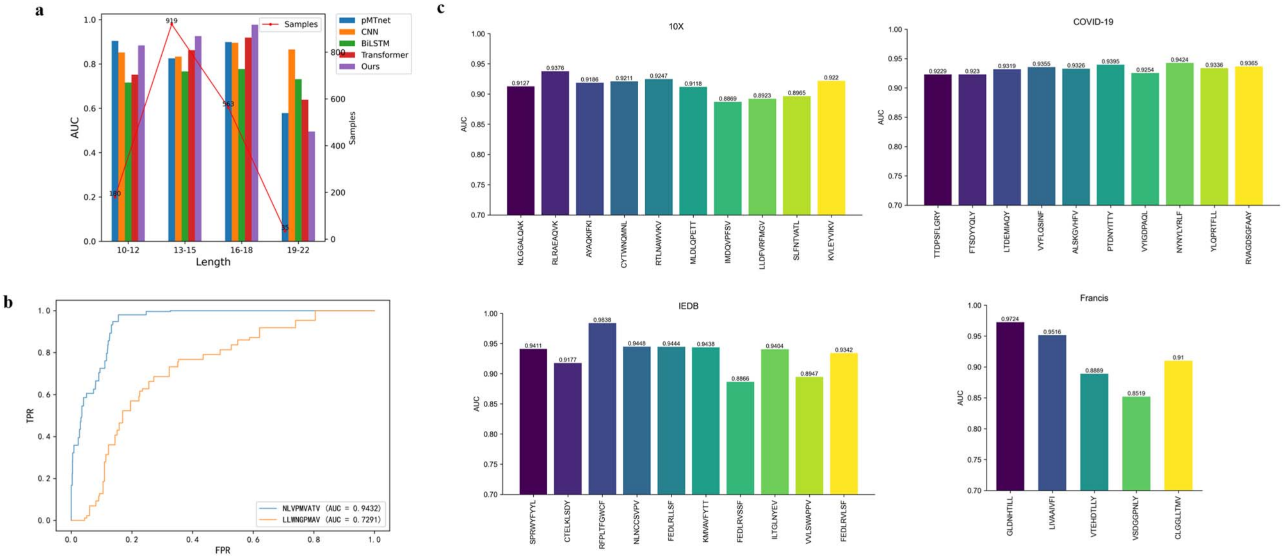


Figure 4. The performance of LightCTL for different TCR sequence lengths and unique peptides. (a) Comparison of the four models under different sequence lengths on the YFV dataset. (b) Results of different antigens in the YFV dataset. (c) Relationship of AUC for unseen epitope in the four independent datasets (i.e. 10X, COVID-19, IEDB, and Francis’s data).

with unknown HLA types, 274 CMV+ subjects and 323 CMV- subjects were selected. To search for CMV-related antigens, we collected seven antigenic epitopes from the IEDB database, namely ELRRKMMYM, IPSINVHHY, NLVPMVATV, QIKVRVDMV, TPRVTGGGGAM, YSEHPTFTSQY, and QIKVRVKMV. We assigned these seven epitopes to each TCR in every TCR repertoire and then predicted each TCR–pMHC binding using LightCTL.

First, we analyzed the relationship between predicted probability and the proportion of unique predicted TCR numbers or the sum of predicted clone frequency (Details are described in Supplementary Method 3.1). Specifically, we screened each TCR repertoire for CMV-specific TCRs using different probability cutoffs and then assessed the proportions of predicted TCR number and their clone frequency. With the expectation that CMV infection would induce specific T cell expansion, we can see that the unique predicted TCR proportion is significantly higher in the CMV+ groups than in the CMV- group with all prediction probability cutoffs (Fig. 5a). This result suggests that our algorithm exhibits significant efficacy when the prediction probability is greater than 0.8. It may be due to the fact that TCR repertoires are highly diversified, with very few TCR clones showing antigenic specificity.

Therefore, we further explored the differences in the average predicted probability for each subject between the CMV+ and CMV- groups (Supplementary Method 3.2). We removed the TCRs shared in the CMV+ versus CMV- groups and calculated the average predicted probabilities of all TCRs in each repertoire. The average prediction probability of the CMV+ group is significantly higher than that of the CMV- group when the cutoff of the predicted probability is 0.5, which indicates that LightCTL is able to sort out specific TCRs with CMV relevance from a massive number of TCRs (Fig. 5b). Finally, we analyzed the differences in TCR diversity and the sum of clone frequencies of samples between the CMV+ and CMV- groups under the condition of setting both the predicted probability and the clone frequency cutoff (Supplementary Method 3.3). For this purpose, we screened samples with a TCR prediction probability ≥0.5 and a clone frequency of TCRs >0.1% and analyzed the number of predicted unique

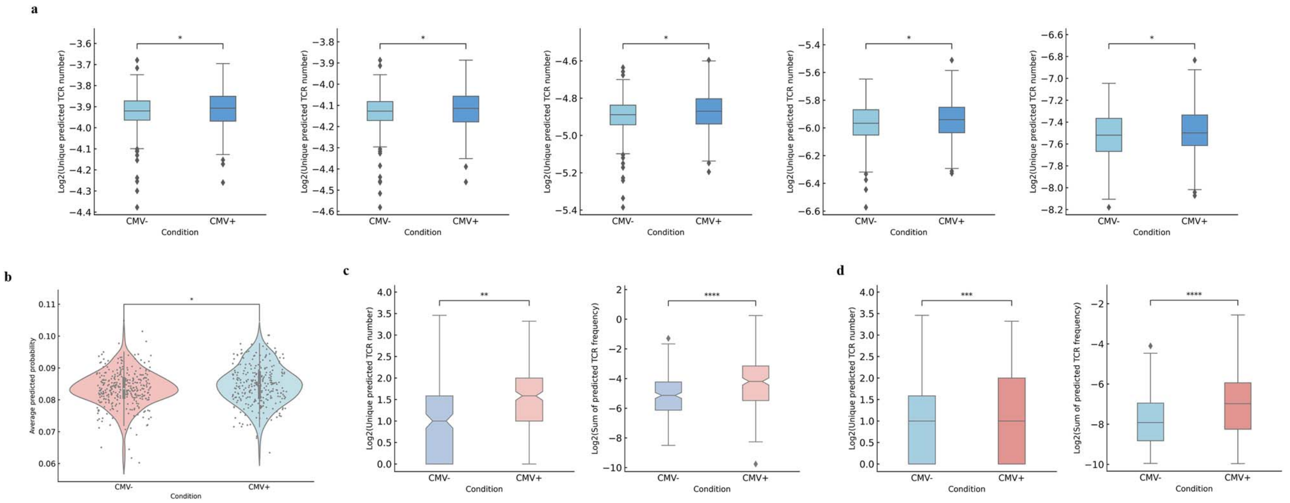


Figure 5. The analysis results of LightCTL in virus immune recognition. (a) Results of unique predicted TCR numbers on CMV+ and CMV- groups when the cutoff is 0.8, 0.9, 0.99, 0.999, and 0.9999, respectively. (b) Comparison of significant differences between average predicted probabilities of TCR in TCR repertoires between CMV+ and CMV- groups when the cutoff of predicted probability is 0.5. (c) Unique predicted TCR number and sum of predicted TCR frequency between CMV+ and CMV- groups on all CMV-associated epitopes under $C_P = 0.5$ and $C_F = 0.001$ (C_P is the cutoff of the TCR prediction probability and C_F is the cutoff of the TCR clone frequency). (d) Unique predicted TCR number and sum of predicted TCR frequency between CMV+ and CMV- groups on CMV-associated specific antigenic epitopes (pp65) under $C_P = 0.5$ and $C_F = 0.001$.

TCRs and the total predicted clone frequency of the samples between the CMV+ and CMV- groups. Both the number of unique TCRs and the total clone frequency of TCRs were significantly higher in the CMV+ group than in the CMV- group, indicating that the results predicted by LightCTL were consistent with the expectation of the virus-induced T cell response (Fig. 5c). Besides, we compared the estimated TCR diversity and clone frequency with a common CMV antigen (pp65) and observed a significant increase in the TCR diversity and clone frequency in CMV+ individuals compared with CMV- individuals (Fig. 5d).

Discussion and conclusion

TCRs trigger T cell activation by specifically binding to antigenic peptides presented by MHC molecules. This process is regulated by MHC restriction, ensuring that T cells respond only to their cognate antigens, thereby mediating an adaptive immune response. Therefore, predicting TCR–pMHC binding has significant biological implications. In basic immunology research, explaining the relationship between the diversity of the TCR library and the breadth of immune response by predicting the binding potential of different TCR clonotypes to pathogen/tumor antigens is of note. In disease diagnosis, therapy, and vaccine design, the prediction of TCR–pMHC binding can help guide the development of bispecific TCRs to target tumor antigens, avoiding off-target toxicity and improving efficacy. However, the high-throughput identification of antigen-specific TCRs from the incredible diversity of TCR pools using a wet-lab experiment remains a significant challenge. Deep learning-based prediction of TCR antigen binding has also encountered difficulties due to the lack of labeled data and the presence of high sequence similarity between positive and negative samples.

LightCTL employs representational learning and attention mechanisms to extract valuable information from TCR sequence, epitope sequence, and MHC subtype, enabling it to predict TCR–pMHC binding with high generalization ability. By leveraging the powerful generalization capabilities of contrastive learning, LightCTL achieves the SOTA performance in predicting TCR-antigen binding specificity. More analyses of the potential application of LightCTL were also performed, including the identification of antigenic specificity under different TCR sequence lengths and the identification of TCR sequence binding specificity under a single antigen sequence. In addition, we analyzed the diversity of CMV-targeted TCR repertoires using TCR bulk sequencing data as a demo case to determine whether the predicted pMHC binding TCRs were indeed expanded in CMV-positive individuals. We also explored potential key sites for TCR epitope binding on a number of TCR–pMHC complexes.

Although LightCTL may have higher generalization ability than other methods, it could still be improved for application in real-world clinical scenarios. Currently, there is limited information on HLA types and epitopes, limiting the downstream application of LightCTL. However, LightCTL can be combined with disease-associated antigen prediction methods [51, 52] and peptide-MHC binding prediction methods [53, 54] to predict more antigen-specific TCRs, and it will be updated as more information on HLAs and epitopes becomes available. It is worth noting that in this study, we did not consider CD4+ T cells in our model, like most previous studies, due to the limited availability of CD4+ T cell data in existing databases. In the future, we will incorporate both CD4+ and CD8+ data by collecting publicly available CD4+ data, generating private data from wet-lab experiments, and developing more generalized methodologies that can consider both CD4+ and CD8+ data. Furthermore, although we trained and evaluated LightCTL using experimentally validated data, and introduced clinically collected and publicly available CMV bulk sequencing data to validate its reliability, its performance may warrant more wet-experiment validations. In the future, the utility of LightCTL will be further explored with more clinical applications.

Methods

Design of LightCTL

A novel lightweight contrastive TCR–pMHC learning framework is proposed to predict the TCR–pMHC binding specificity by using TCR, peptide, and MHC class I (Fig. 1b). First, a TCR encoding module and a pMHC encoding module are designed to extract learnable numeric embedding of TCR and pMHC sequences and extract their shallow features, respectively (Fig. 1c). Secondly, we propose contrastive TCR–pMHC learning paradigms to learn the binding pattern of TCR–pMHC and peptide-MHC. Differing from prior self-supervised contrastive learning methods, contrastive TCR–pMHC learning can learn implicit feature representation by learning the matching relationships between TCRs and pMHCs, so it pays more attention to solving the generalization problem by considering the diversity of sequences. Next, the embedded TCR and pMHC features are fusion and then input into a context-aware prompt module (CAPM) to extract more global feature information (Fig. 1d). The CAPM considers the importance of different amino acid groups in the TCR and pMHC sequences for TCR–pMHC binding specificity prediction. Finally, the prediction results are calculated by an MLP.

TCR encoding module

We consider the input features as single-channel images and design a TCR encoding module and a pMHC encoding module to learn their shallow features (Fig. 1c). The TCR encoding module comprises two convolutional layers, each with a convolutional block, Batch-Normalization, and ReLU. The number of convolutional block channels in the first and second convolutional layers is 16 and 32, respectively. Kernel size, stride, and padding of all convolutional blocks are 3×3, 1, and 1, respectively.

Given a numerical TCR feature $T = [t_1, t_2, t_3, ..., t_n]_{t_n \in R^{1\times L_t \times D_t}}$, the feature map j in the convolutional layer i is

$$F_j^l = \mathbf{Re}LU\{BN[Conv2(T, K_j^l)]\} \tag{1}$$

k_{ij}^l is the weight of the channel j and dimension i in the convolutional layer $l - th$.

pMHC encoding module

Given a peptide embedding $P = [p_1, p_2, p_3, ..., p_n]_{t_n \in R^{1\times L_p \times D_p}}$ and an MHC embedding $M = [m_1, m_2, m_3, ..., m_n]_{t_n \in R^{1\times L_m \times D_m}}$, we first concat the two embeddings to obtain peptide-MHC feature, which is||

$$PM = P||M \tag{2}$$

refers to features of pMHC and TCR concatenated in the third dimension.

Then, we perform a convolution operation on the peptide-MHC feature. We also have two convolutional layers, including convolutional blocks, Batch-Normalization, MaxPooling, and ReLU. The parameter of the pMHC encoding module is the same as the TCR encoding module. Besides, the filter of MaxPooling is 2 × 2.

Contrastive TCR–pMHC learning framework

To improve further the robustness of the model for unseen TCR–pMHC binding specificity prediction, we proposed a contrastive TCR–pMHC learning to enhance the model's expressive capabilities. It is inspired by Contrastive Language-Image Pre-Training [55], whose core idea is to establish a correlation between a TCR and a pMHC and to achieve a better mutual understanding between them through a measure of similarity between the TCR and the pMHC. The method is the first for predicting the TCR–pMHC binding specificity. It differs from existing contrastive learning methods for the prediction of TCR–pMHC binding specificity in that it takes complete account of feature information across data sources between TCRs and pMHCs, and antigen peptides and MHCs, thus improving the predictive power of unknown TCR-antigen binding specificity.

The contrastive TCR–pMHC learning has two parts: CL_{T2PM} and CL_{P2M}. They are trained to map the input into the same embedding space. The goal of CL_{T2PM} is to make similar TCRs and pMHCs closer together in embedding space. The goal of CL_{P2M} is to make similar peptides and MHCs closer together in embedding space. CL_{P2M} is

$$CL_{P2M} = -\log \frac{\mathbf{exp}(sim(K(z_p)_i, V(z_p)_j)/\tau)}{\sum_{k=1}^{2N} 1_{[k\neq i]}\,\mathbf{exp}(sim(K(z_p)_i, V(z_p)_k)/\tau)} \tag{3}$$

And CL_{T2PM} can be calculated by

$$\mathrm{CL}_{\mathrm{T2PM}} = \frac{1}{2}(CL_{TCR2pMHC} + CL_{pMHC2TCR}); \tag{4}$$

$$CL_{TCR2pMHC} = -\log \frac{\mathbf{exp}(sim(z_t, z_p)/\tau)}{\sum_{k=1}^{2N} 1_{[k\neq p]}\,\mathbf{exp}(sim(z_i, z_p)/\tau)}; \tag{5}$$

$$CL_{pMHC2TCR} = -\log \frac{\mathbf{exp}(sim(z_p, z_t)/\tau)}{\sum_{k=1}^{2N} 1_{[k\neq t]}\,\mathbf{exp}(sim(z_p, z_t)/\tau)}, \tag{6}$$

where $1_{[\bullet]} \in \{0, 1\}$ is an indicator function evaluating to 1 if $k = i$ and τ denotes a temperature parameter. $sim(\bullet)$ is cosine similarity. z is the embedded feature for each sequence. $K(\bullet)$ and $V(\bullet)$ are data transformation randomly.

Context-aware prompt module

As specific amino acid motifs are associated with T cell activation, antigen recognition, and particular diseases (cancer, autoimmune diseases, etc.), the interrelationships between amino acid groups may contain information leading to a more in-depth understanding of the mode of interaction between T cells and antigenic peptides. To extract amino acid motifs in the TCR weighing more in T cell activation, we designed a CAPM to enhance the adaptive weighting of the extracted amino acid motif features (Fig. 1d). We first fused the shallow TCR, peptide-MHC information extracted by the TCR encoding module, and pMHC encoding module to obtain the inputs for the CAPM. Given an encoded TCR $z_t \in R^{C\times H\times W}$ and pMHC $z_p \in R^{C\times H\times W}$, their fusion feature z_{tp} is $z_{tp} \in R^{C\times 2H\times W}$. First, a binding feature extraction (BFE) block is used to learn the binding feature, which is

$$z'_{tp} = [\mathbf{Re}LU(BN(Conv(z_{tp})))]_{\times 3} \tag{7}$$

The comprehensive channel feature $z_c \in R^{C\times 1\times 1}$ can be calculated by

$$z_c = AvgPool(z'_{tp}) \tag{8}$$

Then, based on the Squeeze-and-Excitation (SE) block, the weights of the different feature maps of the attention layer i can be calculated by

$$w^i = \frac{1}{1 + e^{-\mathrm{SE}(z_c)}}. \tag{9}$$

Finally, the output of the channel attention layer i is obtained by

$$O_{att}^i = z'_{tp} \times w^i. \tag{10}$$

CAPM is a CNN-based architecture that centers on performing convolution operations on the input data using sliding convolution kernels (filters). Each convolution kernel focuses on local regions of the combined TCR and pMHC feature maps, allowing it to learn their local features (detailed information). CAPM has two attention layers, each containing three BEFs, which aggregate features from the local regions so that features from different layers can be combined. In this way, the later layers can capture more global features and patterns.

Key Points

- A lightweight contrastive TCR–pMHC learning (LightCTL) was proposed to model the complex biological process of TCR-specific recognition of antigens by learning the binding patterns of TCR and antigenic peptides presented by MHC, as well as the binding patterns of MHC and antigenic peptides.
- A CAPM was designed to mine potential feature information related to T cell activation, antigen recognition, and specific diseases by considering the importance of the different feature maps extracted.
- LightCTL achieves superior performance, including accuracy and generalization, compared with previous work in the field and facilitates TCR-related applications.

Supplementary data

Supplementary data is available at *Briefings in Bioinformatics* online.

Conflict of interest: None declared.

Funding

This work is supported by Scientific Research Start-up Funds (QD2021005N), Shenzhen Science and Technology Program (WDZC20220819134430002) and Shenzhen Universities Stable Funding Key Projects (WDZC20200821104802001).

Data and code availability statements

All publicly available datasets and Source codes in this study are available at https://github.com/YYYYYeFei/LightCTL.git.

References

1. Shang J, Jiao Q, Chen C. *et al.* Pretraining transformers for TCR-pMHC binding prediction. In: *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Las Vegas, NV, USA: IEEE, 2022, pp. 26–31, https://doi.org/10.1109/BIBM55620.2022.9994875
2. Van Rhijn I, Branch D, Moody. Cd1 and mycobacterial lipids activate human T cells. *Immunol Rev* 2015;**264**:138–53. https://doi.org/10.1111/imr.12253
3. Murugan A, Mora T, Walczak AM. *et al.* Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci* 2012;**109**:16161–6. https://doi.org/10.1073/pnas.1212755109
4. Croce G, Bobisse S, Moreno DL. *et al.* Deep learning predictions of TCR-epitope interactions reveal epitope-specific chains in dual alpha T cells. *Nat Commun* 2024;**15**:3211. https://doi.org/10.1038/s41467-024-47461-8
5. Wooldridge L, Ekeruche-Makinde J, van den Berg HA. *et al.* A single autoimmune T cell receptor recognizes more than a million different peptides. *J Biol Chem* 2012;**287**:1168–77. https://doi.org/10.1074/jbc.M111.289488
6. Rossjohn J, Gras S, Miles JJ. *et al.* T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol* 2015;**33**:169–200. https://doi.org/10.1146/annurev-immunol-032414-112334
7. Altman JD, Moss PAH, Goulder PJR. *et al.* Phenotypic analysis of antigen-specific t lymphocytes. *Science* 1996;**274**:94–6. https://doi.org/10.1126/science.274.5284.94
8. Zhang S-Q, Ma K-Y, Schonnesen AA. *et al.* High-throughput determination of the antigen specificities of T cell receptors in single cells. *Nat Biotechnol* 2018;**36**:1156–9. https://doi.org/10.1038/nbt.4282
9. Kula T, Dezfulian MH, Wang CI. *et al.* T-Scan: a genome-wide method for the systematic discovery of T cell epitopes. *Cell* 2019;**178**:1016–1028.e13. https://doi.org/10.1016/j.cell.2019.07.009
10. Goncharov M, Bagaev D, Shcherbinin D. *et al.* VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2. *Nat Methods* 2022;**19**:1017–9. https://doi.org/10.1038/s41592-022-01578-0
11. Tickotsky N, Sagiv T, Prilusky J. *et al.* McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences. *Bioinformatics* 2017;**33**:2924–9. https://doi.org/10.1093/bioinformatics/btx286
12. Vita R, Blazeska N, Daniel M. *et al.* The immune epitope database (IEDB): 2024 update. *Nucleic Acids Res* 2025;**53**:D436–43. https://doi.org/10.1093/nar/gkae1092
13. Zhang W, Wang L, Liu K. *et al.* PIRD: Pan immune repertoire database. *Bioinformatics* 2020;**36**:897–903. https://doi.org/10.1093/bioinformatics/btz614
14. Glanville J, Huang H, Nau A. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* 2017;**547**:94–8. https://doi.org/10.1038/nature22976
15. Jokinen E, Huuhtanen J, Mustjoki S. *et al.* Predicting recognition between T cell receptors and epitopes with TCRGP. *PLoS Comput Biol* 2021;**17**:e1008814. https://doi.org/10.1371/journal.pcbi.1008814
16. De Neuter N, Bittremieux W, Beirnaert C. *et al.* On the feasibility of mining cd8+ T cell receptor patterns underlying immunogenic peptide recognition. *Immunogenetics* 2018;**70**:159–68. https://doi.org/10.1007/s00251-017-1023-5
17. Pham M-DN, Nguyen T-N, Tran LS. *et al.* epiTCR: a highly sensitive predictor for TCR–peptide binding. *Bioinformatics* 2023;**39**:btad284. https://doi.org/10.1093/bioinformatics/btad284
18. Zhang Z, Xiong D, Wang X. *et al.* Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. *Nat Methods* 2021;**18**:92–9. https://doi.org/10.1038/s41592-020-01020-3
19. Lin X, George JT, Schafer NP. *et al.* Rapid assessment of T-cell receptor specificity of the immune repertoire. *Nat Comput Sci* 2021;**1**:362–73. https://doi.org/10.1038/s43588-021-00076-1
20. Rollins ZA, Curtis MB, Faller R. *et al.* Automated protein-protein structure prediction of the T cell receptor-peptide major histocompatibility complex. *bioRxiv* 2022;2022.06. 01.494331. https://doi.org/10.1101/2022.06.01.494331
21. Bradley P. Structure-based prediction of T cell receptor: peptide-MHC interactions. *eLife* 2023;**12**:e82813. https://doi.org/10.7554/eLife.82813
22. Gowthaman R, Pierce BG. TCRmodel: high resolution modeling of T cell receptors from sequence. *Nucleic Acids Res* 2018;**46**:W396–401. https://doi.org/10.1093/nar/gky432

23. Yin R, Ribeiro-Filho HV, Lin V. *et al.* TCRmodel2: high-resolution modeling of T cell receptor recognition using deep learning. *Nucleic Acids Res* 2023;**51**:W569–76. https://doi.org/10.1093/nar/gkad356
24. Zhao Y, He B, Li C. *et al.* DeepAIR: a deep-learning framework for effective integration of sequence and 3D structure to enable adaptive immune receptor analysis. *Science Advances* 2023;**9**:eabo5128.
25. Peng X, Lei Y, Feng P. *et al.* Characterizing the interaction conformation between T-cell receptors and epitopes with deep learning. *Nat Mach Intell* 2023;**5**:395–407. https://doi.org/10.1038/s42256-023-00634-4
26. Jurtz VI, Jessen LE, Bentzen AK. *et al.* NetTCR: sequence-based prediction of TCR binding to peptide-MHC complexes using convolutional neural networks. *bioRxiv* 2018;433706. https://doi.org/10.1101/433706
27. Tianshi L, Zhang Z, Zhu J. *et al.* Deep learning-based prediction of the T cell receptor–antigen binding specificity. *Nature. Mach Intell* 2021;**3**:864–75. https://doi.org/10.1038/s42256-021-00383-2
28. Weber A, Born J, Martínez MR. TITAN: T-cell receptor specificity prediction with bimodal attention networks. *Bioinformatics* 2021;**37**:i237–44. https://doi.org/10.1093/bioinformatics/btab294
29. Xu Z, Luo M, Lin W. *et al.* DLpTCR: an ensemble deep learning framework for predicting immunogenic peptide recognized by T cell receptor. *Brief Bioinform* 2021;**22**:bbab335. https://doi.org/10.1093/bib/bbab335
30. Ying X, Qian X, Tong Y. *et al.* AttnTAP: a dual-input framework incorporating the attention mechanism for accurately predicting TCR-peptide binding. *Front Genet* 2022;**13**:942491.
31. Darmawan JT, Leu J-S, Avian C. *et al.* MITNet: a fusion transformer and convolutional neural network architecture approach for T-cell epitope prediction. *Brief Bioinform* 2023;**24**:bbad202. https://doi.org/10.1093/bib/bbad202
32. Jie W, Qi M, Zhang F. *et al.* TPBTE: a model based on convolutional transformer for predicting the binding of TCR to epitope. *Mol Immunol* 2023;**157**:30–41.
33. Motuzenko K, Makarov I. Analyzing immunomes using sequence embedding and network analysis. In: *2023 IEEE 21st World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 000325–30. Herl'any, Slovakia: IEEE, 2023, pp. 000325–30. https://doi.org/10.1109/SAMI58000.2023.10044509
34. Zhao Y, Xiaona S, Zhang W. *et al.* SC-AIR-BERT: a pre-trained single-cell model for predicting the antigen-binding specificity of the adaptive immune receptor. *Brief Bioinform* 2023;**24**:bbad191. https://doi.org/10.1093/bib/bbad191
35. Fang Y, Liu X, Liu H. Attention-aware contrastive learning for predicting T cell receptor–antigen binding specificity. *Brief Bioinform* 2022;**23**:bbac378. https://doi.org/10.1093/bib/bbac378
36. Gao Y, Gao Y, Fan Y. *et al.* Pan-peptide meta learning for T-cell receptor–antigen binding recognition. *Nat Mach Intell* 2023;**5**:236–49. https://doi.org/10.1038/s42256-023-00619-3
37. Xiao Y, Yueshan Huang Y, Zhao FX. *et al.* Multimodal-AIR-BERT: a multimodal pre-trained model for antigen specificity prediction in adaptive immune receptors. In: *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, Istanbul, Turkey: IEEE, 2023, pp. 69–75. https://doi.org/10.1109/BIBM58861.2023.10385479
38. Nielsen M, Lundegaard C, Blicher T. *et al.* NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and-B locus protein of known sequence. *PloS One* 2007;**2**:e796. https://doi.org/10.1371/journal.pone.0000796
39. Pogorelyy MV, Minervina AA, Touzel MP. *et al.* Precise tracking of vaccine-responding T cell clones reveals convergent and personalized response in identical twins. *Proc Natl Acad Sci* 2018;**115**:12704–9. https://doi.org/10.1073/pnas.1809642115
40. 10X_Genomics. A new way of exploring immunity: linking highly multiplexed antigen recognition to immune repertoire and phenotype (application note lit000047 rev c). retrieved from 10x genomics website. *Retrieved from the 10X Genomics website*, 2020.
41. Nolan S, Vignali M, Klinger M. *et al.* A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-Cov-2. *Res Sq* 2020;rs. 3. rs-51964.
42. Francis JM, Leistritz-Edwards D, Dunn A. *et al.* Allelic variation in class I HLA determines cd8+ T cell repertoire shape and cross-reactive memory responses to SARS-CoV-2. *Sci Immunol* 2022;**7**:eabk3070.
43. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;**9**:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735
44. Zhang Y, Ye F, Gao X. MCA-Net: multi-feature coding and attention convolutional neural network for predicting lncRNA–disease association. *IEEE/ACM Trans Comput Biol Bioinform* 2021;**19**:2907–19.
45. Vaswani A, Shazeer N, Parmar N. *et al.* Attention is all you need. In: Guyon I, Luxburg UV, Bengio S. *et al.* (eds.), *Advances in Neural Information Processing Systems 30 (NIPS 2017)*. Long Beach, CA, USA: Curran Associates, Inc.; 2017, pp. 6000–6010.
46. Feng Z, Chen J, Hai Y. *et al.* Sliding-attention transformer neural architecture for predicting T cell receptor–antigen–human leucocyte antigen binding. *Nat Mach Intell* 2024;**6**:1216–30. https://doi.org/10.1038/s42256-024-00901-y
47. Zhang J, Ma W, Yao H. Accurate TCR-pMHC interaction prediction using a BERT-based transfer learning method. *Brief Bioinform* 2024;**25**:bbad436. https://doi.org/10.1093/bib/bbad436
48. Zhang Y, Wang Z, Jiang Y. *et al.* Epitope-anchored contrastive transfer learning for paired cd8+ T cell receptor–antigen recognition. *Nat Mach Intell* 2024;**6**:1344–58.
49. Leem J, de Oliveira SHP, Krawczyk K. *et al.* STCRDab: the structural T-cell receptor database. *Nucleic Acids Res* 2018;**46**:D406–12. https://doi.org/10.1093/nar/gkx971
50. Emerson RO, DeWitt WS, Vignali M. *et al.* Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet* 2017;**49**:659–65. https://doi.org/10.1038/ng.3822
51. Tran T-O, Le NQK. Sa-TTCA: an SVM-based approach for tumor T-cell antigen classification using features extracted from biological sequencing and natural language processing. *Comput Biol Med* 2024;**174**:108408. https://doi.org/10.1016/j.compbiomed.2024.108408
52. Huber F, Arnaud M, Stevenson BJ. *et al.* A comprehensive proteogenomic pipeline for neoantigen discovery to advance personalized cancer immunotherapy. *Nat Biotechnol* 2024. https://doi.org/10.1038/s41587-024-02420-y
53. Chang Y, Ligang W. CapHLA: a comprehensive tool to predict peptide presentation and binding to HLA class I and class II. *Brief Bioinform* 2025;**26**:bbae595.

54. Marzella DF, Crocioni G, Radusinović T. *et al.* Geometric deep learning improves generalizability of MHC-bound peptide predictions. *Commun Biol* 2024;**7**:1661. https://doi.org/10.1038/s42003-024-07292-1

55. Radford A, Kim JW, Hallacy C. *et al.* Learning transferable visual models from natural language supervision. In: Meila M, Zhang T (eds.), *Proceedings of the 38th International Conference on Machine Learning (ICML 2021)*. Virtual: PMLR; 2021, pp. 8748–63.

Briefings in Bioinformatics, 2025, **26(3)**, bbaf246
https://doi.org/10.1093/bib/bbaf246
Problem Solving Protocol