IMMUNOTHERAPY

Improved MHC II epitope prediction — a step towards personalized medicine

Tamson V. Moore and Michael I. Nishimura

Numerous neoepitope-based vaccination strategies are in testing for clinical use in the treatment of cancer. Rapid identification of immunostimulatory neoantigen targets hastens neoantigen vaccine development. Papers recently published in *Nature Biotechnology* describe two independent machine-learning-based algorithms that demonstrate improved identification of MHC class II-binding peptides. Herein, we outline the benefits of these algorithms and their implications for future immunotherapies.

Refers to Racle, J. et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* **37**, 1283–1286 (2019) | Chen, B. et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat. Biotechnol.* **37**, 1332–1343 (2019).

Immunotherapies, including immune-checkpoint inhibitors (ICIs), can induce durable tumour regression and even disease remission in a diverse subset of patients with chemotherapy-refractory metastatic cancers. The efficacy of ICIs is generally greater in cancer types with higher median numbers of somatic mutations[1], which can generate neoantigens that are targets for specific CD4[+] and/or CD8[+] T cells[2]. Evidence indicates that CD4[+] T cell responses to MHC class II (MHC II)-restricted antigens are required for robust responses to ICIs[3] and that neoantigen vaccines can enhance CD4[+] T cell responses[2]. To develop effective neoantigen vaccines, it is essential to identify neoantigen epitopes (neoepitopes) that will bind to MHC II molecules and be presented to CD4[+] T cells. Whereas the presence and expression of neoantigen proteins can be identified through sequencing of the tumour exome, the neoepitopes presented by MHC II molecules must be either discovered empirically using expensive and time-consuming mass spectrometry (MS) techniques[4] or predicted using software-based estimations of peptide–MHC II binding affinity.

In the November issue of *Nature Biotechnology*, the authors of two independent studies[5,6] described novel machine-learning algorithms for identifying MHC II-binding peptides. Chen et al.[6] developed the MHC analysis with recurrent integrated architecture (MARIA) platform, in which neural network-based models trained on large MS-based peptide datasets are used to generate a peptide presentation score, given inputs of a query peptide sequence and corresponding gene name in addition to MHC II (*HLA-D*) alleles. Racle et al.[5] developed MoDec, a motif deconvolution algorithm with conceptual similarity to convolutional neural networks, to identify MHC II-binding motifs, binding core offset preferences and peptide cleavage motifs from large MS-based peptidome datasets encompassing HLA-DR, HLA-DQ and HLA-DP alleles. The deconvoluted peptidomic datasets were then used to train a prediction algorithm, MixMHC2pred, which returns an MHC II binding score for a given peptide sequence and *HLA-D* allele. When tested on known MHC II-binding epitopes and decoy epitopes, both the MARIA and MixMHC2pred algorithms had significantly improved predictive accuracy compared with

> ❝ MARIA and MixMHC2pred algorithms had significantly improved predictive accuracy… ❞

SMN Align ($P < 1 \times 10^{-5}$) and NetMHCIIpan ($P < 0.001$), respectively[5,6], which are two commonly used MHC II-binding prediction algorithms.

In both studies[5,6], large (approximately 50,000–100,000 peptides) MS-based datasets of MHC II-presented peptides were used to train independent algorithms to estimate peptide–MHC II binding. However, the algorithms have different advantages. For example, MARIA incorporates tissue-specific gene-expression levels, in order to account for effects of the abundance of protein on the likelihood of peptide presentation by MHC II molecules, whereas MixMHC2pred does not. MixMHC2pred, with MoDec, used a larger training dataset (~100,000 peptides compared with ~50,000 for MARIA), encompassing more cell types, and enables the identification of peptides that bind to the different MHC II isotypes (encoded by the *HLA-DR*, *HLA-DP* and *HLA-DQ* genes), without retraining, whereas the MARIA benchmarks were established using versions of the algorithm trained independently for different MHC II isotypes.

The large datasets used to train these algorithms improved both the accuracy and specificity of MHC II-binding predictions[5,6]. MS is becoming increasingly popular as a method of identifying the peptidome from a variety of tumour types[7] and the resultant increased dataset availability for the training of MHC II-binding algorithms will probably further improve the accuracy of these algorithms over time. Nonetheless, a key caveat of using software-based modelling instead of empirical testing is the inability to identify outliers — MHC II-binding algorithms model the average ways in which most peptides bind (thus identifying recurrent motifs) and are likely to exclude peptides that bind MHC II molecules in unusual ways.

With regard to the clinical goal of predicting CD4[+] T cell reactivity, MixMHC2pred identified a higher number of true immunogenic MHC II-binding epitopes than NetMHCIIpan, as demonstrated in vitro using CD4[+] T cells

isolated from two patients with melanoma[5]. Similarly, MARIA successfully identified patient-specific neo epitopes with reactive CD4[+] T cells in two of three patients with mantle cell lymphoma[6]. Although these datasets are small, they indicate that both algorithms can accurately predict peptides that can stimulate CD4[+] T cells. One remaining hurdle, however, is that only a minority of predicted MHC II-binding peptides induced CD4[+] T cell responses (8.3% (5 of 60) with MixMHC2pred and 10.8% (20 of 185) with MARIA)[5,6]. Notably, in the majority of previous studies, <5% of potential neoepitopes were found to stimulate T cells, even after preselection for MHC binding[8]. However, the absence of a T cell response should not be automatically attributed to the production of false-positive predictions by an algorithm. MHC II-binding peptides can fail to activate T cell responses for several reasons. First, the development of immunity to an MHC II-bound antigen is dependent on the presence of T cells bearing a cognate T cell receptor (TCR). T cells are able to recognize a large pool of antigens[9], made more numerous by cross-reactivity[10]; however, with the 20 proteinogenic amino acids, 4.10×10^{15}–3.28×10^{19} peptides comprising 12–15 amino acid residues (MHC II-binding peptides can contain 9–25 residues) could potentially exist, exceeding the number of T cells in the human body (estimated to be $<10^{13}$). Therefore, few or no reactive T cells might exist for some peptides, resulting in the absence of detectable CD4[+] responses. Second, the assays used to test peptide immunogenicity usually involve several million T cells, at most, meaning that T cell clones with a very low abundance might not be represented or their activity might not be detectable above background levels. Finally, neoantigen-specific T_{reg} cells have been detected in patients with cancer[11], and these cells could potentially suppress the activity of and thus prevent the detection of neoantigen-specific effector CD4[+] T cells in the typical enzyme-linked immunosorbent spot (ELISPOT) immunogenicity assays. Therefore, factors other than MHC II-binding

> **Both MARIA and MixMHC2pred have the potential to make personalized neoantigen-based therapies more accessible to patients…**

might dictate CD4[+] T cell responses to predicted MHC II-binding peptides, including — but not limited to — deficits in the TCR repertoire or suppression of T cells.

MARIA and MixMHC2pred both enabled enhanced detection of CD4[+] T cell-stimulating neoantigen peptides and reduced false-positive rates compared with prior platforms. Therefore, both algorithms are usefully improved tools for identifying MHC II-binding neoepitopes, as long as the low rate of CD4[+] T cell response is taken into account and a sufficient number of peptides to induce a response are included in any experimental vaccines. Historically, neoepitope-based vaccines have demonstrated clinical benefit as single agents, mostly in the adjuvant or prophylactic setting[2]. The limited efficacy of cancer vaccines in the treatment of unresectable metastatic disease has been largely attributed to tumour-mediated immunosuppression. Combinations of neoepitope vaccines with ICIs that reduce immunosuppression have, however, shown promise in the treatment of non-resected aggressive cancers in mice[12]; this combination strategy is currently being tested in multiple clinical trials (for example, NCT03532217, NCT03568058, NCT03639714, NCT03970382 and NCT03597282). Neoepitope vaccines could potentially also be combined with adoptive T cell therapies. Specifically, neoepitope vaccination of patients is being used to promote the expansion of neoepitope-specific T cells in order to facilitate the cloning of patient-specific neoepitope-specific TCRs, with subsequent ex vivo genetic modification of large numbers of autologous non-tumour-reactive T cells to express the neoepitope-specific TCRs before

they are returned to the patient to attack the tumour (NCT03412877 and NCT03970382).

The efficacy of such neoantigen-based immunotherapies will be dependent on the identification of a sufficient number of MHC II-binding peptides to stimulate CD4[+] T cell responses. Both MARIA and MixMHC2pred have the potential to make personalized neoantigen-based therapies more accessible to patients, including patients with tumours harbouring fewer mutations, by identifying more MHC II-binding epitopes to which CD4[+] T cells can respond within each patient's pool of putative neoantigens.

Tamson V. Moore and Michael I. Nishimura*

Department of Surgery, Loyola University Chicago, Maywood, IL, USA.

*e-mail: mnishimura@luc.edu

1. Keenan, T. E., Burke, K. P. & Van Allen, E. M. Genomic correlates of response to immune checkpoint blockade. *Nat. Med.* **25**, 389–402 (2019).
2. Ott, P. A. et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).
3. Alspach, E. et al. MHC-II neoantigens shape tumour immunity and response to immunotherapy. *Nature* **574**, 696–701 (2019).
4. Bassani-Sternberg, M. et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **7**, 13404 (2016).
5. Racle, J. et al. Robust prediction of HLA class II epitopes by deep motif deconvolution of immunopeptidomes. *Nat. Biotechnol.* **37**, 1283–1286 (2019).
6. Chen, B. et al. Predicting HLA class II antigen presentation through integrated deep learning. *Nat. Biotechnol.* **37**, 1332–1343 (2019).
7. Zhang, X. et al. Application of mass spectrometry-based MHC immunopeptidome profiling in neoantigen identification for tumor immunotherapy. *Biomed. Pharmacother.* **120**, 109542 (2019).
8. Karpanen, T. & Olweus, J. The potential of donor T-cell repertoires in neoantigen-targeted cancer immunotherapy. *Front. Immunol.* **8**, 1718 (2017).
9. Arstila, T. P. et al. A direct estimate of the human alphabeta T cell receptor diversity. *Science* **286**, 958–961 (1999).
10. Kudela, P. et al. Cross-reactive CD4⁺T cells against one immunodominant tumor-derived epitope in melanoma patients. *J. Immunol.* **179**, 7932–7940 (2007).
11. Ahmadzadeh, M. et al. Tumor-infiltrating human CD4+ regulatory T cells display a distinct TCR repertoire and exhibit tumor and neoantigen reactivity. *Sci. Immunol.* **4**, eaao4310 (2019).
12. Tondini, E. et al. A poly-neoantigen DNA vaccine synergizes with PD-1 blockade to induce T cell-mediated tumor control. *Oncoimmunology* **8**, 1652539 (2019).

Competing interests
The authors declare no competing interests.