HABIT – a webserver for interactive T cell neoepitope discovery

Joana Martins[1,2], Carlos Magalhães[1,2], Vítor Vieira[3], Miguel Rocha[3] and Nuno S. Osório[1,2]

1- Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal. 2- ICVS/3B PT Government Associate Laboratory, Braga/Guimarães, Portugal. 3- Centre of Biological Engineering, University of Minho, Braga, Portugal.

Summary:

Neoepitopes generated by amino acid variants specifically found in pathogens or cancer cells are gaining momentum in immunotherapy development. HABIT (HLA Binding InTelligence) is a web platform designed to generate and analyse machine learning-based T cell epitope predictions for improved neoepitope discovery.

Availability and Implementation:

HABIT is available at http://habit.evobiomed.com. Peptide-HLA binding predictions by NetMHCpan 4.0 and NetMHCIIpan 3.1 were implemented in a web application for interactive data exploration using shiny package powered by RStudio.

Contact: nosorio@med.uminho.pt

Supplementary information: Supplementary data and video tutorials are available online.

1 Introduction

Immune responses mediated by T cells are triggered by specific peptides (T cell epitopes) bound to the highly diverse human leukocyte antigens (HLA) (Blum *et al*., 2013). Variations in one single amino acid might impact peptide-HLA interactions and subsequent T cell responses (Unanue *et al*., 2016; Hilf *et al*., 2018). The machine learning-based methods NetMHCpan (Vanessa *et al*., 2017) and NetMHCIIpan (Andreatta *et al*., 2015) rank among the best overall predictors of peptide-HLA binding (Trolle *et al*, 2015, Andreatta *et al*., 2017). However, using this type of tools to identify neoepitopes, such as the ones generated by cancer or pathogen-specific amino acid mutations, is not trivial. It requires individual prediction of wild-type and mutant sequences, the selection of all overlapping peptides including the amino acid positions of interest and the comparison of the results generated separately. Thus, programming skills are needed to extract meaningful conclusions from the large outputs of HLA-peptide binding prediction data. Furthermore, data visualization and prioritization strategies, such as statistical tests or human population coverage calculation are inexistent in an integrated manner and would be of great value for immune-based

therapy and diagnostic strategies (Oyarzun *et al.*, 2015). Taking this into account, we developed HABIT, which automates HLA binding prediction and provides advanced interpretation of the impact of amino acid variants in peptide-HLA binding (class I and II). It features an intuitive, user-friendly and interactive graphical interface that plots, summarizes and performs statistical comparisons on the antigenicity scores of wild-type and mutant epitope candidates.

2 Implementation

HABIT was developed in the RStudio v3.5.0, using the shiny v. 3.5.0 package to create the web-based application. Plotly v.4.8.0 and DataTables v.0.4 packages were used for the generation of interactive plots and customizable tables. NetMHCpan 4.0 and NetMHCIIpan 3.1 command line tools available for Linux were implemented in the back-end of HABIT to predict the peptide fragments with high binding affinity to HLA class I and class II molecules. All 8 to 14-mer peptides for wild-type and mutant sequences overlapping the position of the mutation of interest can be predicted to a set of 2924 class I alleles. Simultaneously, NetMHCIIpan can predict the binding affinity of 9-30 mer peptides to 660 HLA class II alleles. Predicted antigenic scores are plotted and summarized enabling efficient selection of neoepitope candidates.

3 Results

To illustrate the relevance of the features implemented in HABIT, we analysed the effect of a somatic mutation (TP53 R248W) found in several types of human cancer (Tate *et al.*, 2018). Simultaneous HLA-binding prediction with NetMHCpan and NetMHCIIpan was launched for all wild-type and mutant peptides containing the amino acid residue of interest (Figure 1A and B). An interactive visual representation of the impact the R248W mutation had on HLA binding affinity was obtained. As shown in Figure 1C, the mutant amino acid generated several novel peptides predicted to bind strongly (in orange) to specific HLA class II proteins (Figure 1C). Interactive tables detailing the differences between wild-type and mutant peptides were also generated (Figure 1D), including percentage of strong binders, mean HLA affinity and mean percentile rank, along with statistical analysis. Sort and search functionality highlighted the HLA class II proteins DRB1_0101 and DRB1_0102 as the ones with higher affinity for the majority of the neoepitopes (Figure 1D). Information about HLA distribution across the globe, as a way to evaluate the impact of neoepitopes in the immune responses across distinct human populations was also obtained (Figure 1E).

Overall, HABIT provides a fast and user-friendly way to study neoepitopes, answering the following questions: what is the predicted impact of amino acid mutations in HLA-binding properties?; are the differences statistically significant?; and what percentage/subset of the human population is more likely to respond to the neoepitopes? With this, HABIT might advance the discovery of molecular determinants that influence variation in T cell-mediated immune responses, thus contributing to the design of immune system-based therapy and diagnostic tools.

References

Blum, J. S. *et al*. (2013) Pathways of Antigen Processing. *Annual Review of Immunology* **31**, 443-473.

Hilf, N. *et al*. (2019) Actively personalized vaccination trial for newly diagnosed glioblastoma." *Nature* **565**, 240–245.

Unanue, E. R. *et al*. (2016) Variations in MHC class II antigen processing and presentation in health and disease. *Annual review of immunology* **34**, 265-297.

Vanessa, J. *et al.* (2017) NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *The Journal of Immunology* **199**, 3360–3368

Andreatta, M. *et al.* (2015) Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. *Immunogenetics* **67,** 641–650.

Trolle, T. *et al.* (2015) Automated benchmarking of peptide-MHC class I binding predictions. *Bioinformatics* **31,** 2174–2181.

Andreatta, M. *et al*. (2017) An automated benchmarking platform for MHC class II binding prediction methods. *Bioinformatics* **34**, 1522-1528.

Oyarzun, P. *et al*. (2015) Computer-aided design of T-cell epitope-based vaccines: addressing population coverage. *International Journal of Immunogenetics* **42**, 313-321.

Tate, J. *et al*. (2018) COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research* **47**, 941–947.

Allele	Wild type			Mutant			Wild type vs Mutant			Population coverage
	Predicted SB Epitopes	Mean Binding Affinity (nM)	Mean % Rank	Predicted SB Epitopes	Mean Binding Affinity (nM)	Mean % Rank	% Predicted SB epitopes (Mutant – Wild type)	Mean Binding Affinity (nM) (t-test p-value)	Mean % Rank (t-test p-value)	World
DRB1_0101	112	167.9	39.1	286	83	22.2	46.5	8.54e-09	3.14e-39	11.53
DRB1_0102	269	57.7	39.5	321	50	30.7	13.9	3.12e-01	1.70e-17	3.19

Allele	World	East Asia	North East Asia	South Asia	South East Asia	South West Asia	Europe	East Africa	West Africa	Central Africa	North Africa	South Africa	West Indies	North America	Central America	South America	Oceania
DRB1_0101	11.53	10.21	2.77	9.25	0.78	2.67	15.19	3.96	4.47	2.12	2.33	0	9.55	13.55	1.34	4.05	0.81
DRB1_0102	3.19	0.07	0.37	0	0	4.03	2.12	11.83	6.42	9.63	8.77	0	5.69	4.06	2.11	2.81	0

Figure 1. Features of HABIT webserver. A and B) User-definable parameters: protein and amino acid mutation of interest. C) Antigenicity scores: visual comparison of the HLA-binding affinity (nM) between wild-type (x axis) and mutant (y axis) peptides. Classification of peptides in non binder=green; strong

binder(SB)=orange and weak binder(WB)=purple. D) Summary and statistical analysis of prediction metrics (binding affinity and % rank) for each HLA allele. E) HLA frequency across 17 world geographic regions.