Computational MHC-I epitope predictor identifies 95% of experimentally mapped HIV-1 clade A and D epitopes in a Ugandan cohort.

Daniel Bugembe Lule (dan.lule@mrcuganda.org)
MRC / UVRI & LSHTM Uganda Research Unit https://orcid.org/0000-0002-5616-515X
Andrew Obuku Ekii
MRC/UVRI Uganda Research Unit On Aids
Christine Watera
Uganda Virus Research Institute
Nicaise Ndembi
Institute of Human Virology, Abuja, Nigeria
Jennifer Serwanga
MRC/UVRI & LSHTM Uganda Resaerch Unit
Pontiano Kaleebu
MRC Lifecourse Epidemiology Unit
Pietro Pala
MRC/UVRI Uganda Research Unit On Aids

Research article

Keywords: HIV-1, epitope mapping, T-cell, artificial neural network, in-silico and NetMHCpan4.0

Posted Date: September 16th, 2019

DOI: https://doi.org/10.21203/rs.2.14495/v1

Version of Record: A version of this preprint was published on February 22nd, 2020. See the published version at https://doi.org/10.1186/s12879-020-4876-4.

Abstract

Background: Identifying immunogens that induce HIV-specific immune responses is a lengthy process that can benefit from computational methods, which predict T-cell epitopes for various HLA types.

Methods: We tested the performance of the NetMHCpan4.0 computational neural network in re-identifying 93 T-cell epitopes that had been previously independently mapped using the whole proteome IFN-g ELISPOT assays in 6 HLA class I typed Ugandan individuals infected with HIV-1 subtypes A1 and D.

Results: NetMHCpan4.0 correctly predicted 88 of the 93 experimentally mapped epitopes for a set length of 9-mer and matched HLA class I alleles. Receiver Operator Characteristic (ROC) analysis gave an area under the curve (AUC) of 0.928. Setting NetMHCpan4.0 to predict 11-14mer length did not improve the prediction (37-79 of 93 peptides) with an inverse correlation between the number of predictions and length set. Late time point peptides were significantly stronger binders than early peptides (Wilcoxon signed rank test: p =0.0000005).

Conclusion: NetMHCpan4.0 class I epitope predictions covered 95% of the epitope landscape recognised by HIV-1 infected individuals, and would have reduced the number of experimental confirmatory tests by >80%. Algorithmic epitope prediction in conjunction with HLA allele frequency information can cost-effectively assist immunogen design. Keywords: HIV-1, epitope mapping, T-cell, artificial neural network, in-silico and NetMHCpan4.0.

Background

Computational algorithms are increasingly utilised in biological modelling and offer the potential to reduce the time and expense of immunological assays. Computational algorithms were initially demonstrated as useful tools for predicting potential epitopes that might elicit quality T-cell responses [1, 2]. Computational algorithms that predict potential HLA binding T-cell epitopes can facilitate the design of vaccines capable of inducing T-cell immunity against HIV-1. The high variability of HIV-1 and the extensive genetic polymorphism of HLA molecules can be managed *in silico,* allowing immunogen optimisation to increase breadth and magnitude of T cell responses in respect of HLA allele frequencies and circulating virus strains in different populations. Bioinformatics approaches have been applied as proof of concept for an HIV-1 peptide-based vaccine for the *env* and *gag* genes [3] in cynomolgus macaques for a broad spectrum of HIV-1 clades. Computational optimisation of immunogens facilitates the development of the multivalent and mosaic vaccines [4] necessary to control recombinant HIV-1 strains, an increasingly

common occurrence in the epidemic in Uganda [5]. Computational approaches aim to identify optimal epitopes relevant to vaccine development and are not isolated to HIV only, but a wide range of pathogens, including Ebola virus [6] , consequently some statistical validation approaches have been applied for evaluation of these methods [7-10].

A reliable pan-HLA-specific algorithm NetMHCpan4.0 [11-13] that has been improved by advances in HLA binding data, covers 172 MHC class I molecules from human (HLA-A, B, C, E), mouse (H-2), cattle (BoLA), primates (Patr, Mamu, Gogo) and swine (SLA) [13, 14], and can also predict binding to alleles devoid of experimental data basing on similarity to known binders and non-binders [15, 16]. This is an artificial neural network (ANN) algorithm for predictions of 8-14aa and capable of predicting epitopes for other HLA alleles using data for similar alleles by positional similarity of residues in their binding motifs. The binding of CTL epitopes to MHC class I molecules is linear, anchoring at residues 2 and 9; hence the interface between ligand and CTL can be determined computationally [17]. Validation of such computational applications can be done by comparing their predictions with experimental data. Although computational methods have been reported to achieve an area under the curve AUC of over 90% [11, 12, 18, 19], they had not been evaluated using wet laboratory experimental data. One study that explored the reliability of *in-silico* approaches in epitope prediction and its application for vaccine design reported a meagre 22%, 44%, and relatively higher 78% match for three computational tools namely YFPEITHI, CTLPRED and IEDB respectively [20]. Using experimental epitope mapping data generated from 757 peptides tested on cells of 6 early HIV-1 infected individuals at paired time points, we show that NetMHCpan4.0 can be useful for markedly reducing pooled peptide experiments as demonstrated by the 95% experimental and computational concordance.

Results

Magnitude of Epitope Predictions are Variable across HLA Alleles, HIV-1 Proteins and Clades

The input HIV-1 subtypes A1 and D consensus whole proteome sequences evaluated for potential 9, 10, 11, 12, 13 and 14-mer binders to the 20 HLA alleles represented in the six patients, varied in the distribution of predicted binders across HIV-1 genes and HLA alleles. All the peptide hits predicted for 10 through 14-mer were also all predicted in the 9-mer set except for two 14-mer peptides. An expected positive correlation for protein size with number of predictions as well as with HIV protein length was observed as illustrated by Spearman's rank order correlation; r_s=0.88 (Figures 1, A and B). NetMHCpan4.0 predicted 95% (88/93) of the experimentally mapped peptides as binders and missed 5% (5 out of 93) for the 12-time points of the 6 participants (Table 2). Comparison of the various epitope prediction length set showed that NetMHCpan4.0 predicted 88, 79, 55, 39, 39 and 37 hits out of 93 for 9, 10, 11, 12, 13 and 14-mer epitopes respectively. Increasing the prediction length from 9mer through 14mer resulted in a smaller number of predicted binders as illustrated in Figure 2. The computational predictor however had more predicted binders than those determined by the experimental mapping as presented in the confusion matrix in Table 1. The experimental positive's count also shown in table 2 under column "Hits" shows the test peptide count (1-88) that contained the computational 9-mer sequence. Multiple computational epitopes may be contained in a single experimental peptide, as shown in the column "NetMHCpan4.0 9-mer Epitope Prediction" in Table 2. Overall HIV Clade A 9mer predictions were fewer in number than clade D (figure 1, C and D) though the difference did not approach statistical significance.

Comparison of Experimentally Mapped Epitopes with *in-silico* Prediction

The experimental peptide mapping data was derived from a baseline time point corresponding to HIV Fiebig stages IV, V and VI (table 4) and a later time point. Ninety-three (n=93) epitopes were experimentally mapped of which 12 were recognized at both baseline and later time points, 34 only at baseline and 54 only at the later time point. Comparison of the ranked computational score for Netmhcpan4.0 binders of early (n=34) versus later peptides showed that the later time point predictions were stronger binders reaching statistical significance, determined by the Wilcoxon signed rank test with a *p*-value=0.0000005 (figure 3). NetMHCpan4.0 ranked binders as those predicted to be in the

top 2% and assigned a score of 0.2 or below. Any binder within the top 0.5% and assigned a score of 0.05 or below was ranked as a strong binder.

Considering only the 9-mer computational predictions, peptides that were derived from the same 17-mer experimental peptide were determined by a BLAST mapping to their derivative sequences. The 17-mer peptides were then classified into a confusion matrix (table 1) as either true positives, false positive, true negative and false negative. From the classification the true positive rate (sensitivity) was plotted against the false positive rate (1-specificity) for only 9-mer predictions using an ROC curve and the AUC attained reached 0.928 (Figure 4). Only 9-mer length epitopes were considered in the ROC analysis as increasing the length to 10-mer through 14mer NetMHCpan4.0 predictions neither raised the number of predicted binders nor improved the hit rate as all their predictions contained the sequence already predicted in the 9-mer set except 1 14-mer peptide (hit 72 in table 2). We observed a negative correlation for the number of computationally predicted epitopes with the length of input for netMHCpan4.0 (Figure 2).

Discussion

In this analysis we showed that the computational method NetMHCpan4.0 predicted 95% of previously experimentally mapped HIV-1 epitopes in 6 HIV infected individuals expressing a total of 20 different HLA class I alleles. In our IFN-g ELISPOT assays we evaluated 757 17mer peptides overlapping by 11 amino acids and covering the whole subtype A1 and D consensus proteomes. The NetMHCpan4.0 algorithm scans protein sequences producing binding score outputs for 9-14mer epitopes, therefore, for the purpose of this evaluation, predicted 9-mer epitopes were matched to the experimental 17mer sequences that included them wholly. Out of the 5 experimentally determined epitopes missed by the algorithm, 4 were actually computationally predicted as binders but were not included for lack of concordance with the participant's HLA alleles. About one third (37) of 125 total positive predictions were not experimentally supported in our tests. These do not necessarily represent false positives, as ELISPOT detection depends on the frequency of specific T cells in the participant's repertoire, and we observed changes in dominant T cell specificities within a given participant between early and later time points after HIV-1 infection. A

formal ROC evaluation of the score generated by NetMHCpan4.0 as a classifier for peptides recognised/not recognised by PBMC in IFN-g ELISPOT assays, produced an AUC of 0.928. Thus experimental confirmatory tests cannot be dropped altogether, however the NetMHCpan4.0 algorithm could provide a considerable saving of time and resources in verifying just the predicted epitopes.

We observed a strong correlation between protein size and number of epitopes predicted, with the largest number of epitopes in the Env protein followed by Pol and Gag. Subtype D sequences had more predictions than subtype A1, although the difference was not statistically significant.

As participants had been enrolled in the acute/early phase of HIV-1 infection and we had observed intra-participant changes in epitope recognition between early and late time points after infection, we compared the binding scores of confirmed epitopes at these time points and found a statistically significant change towards recognition of higher binding peptides as the infection entered the chronic phase. This might represent better support of the T-cell response directed at more stable HLA/peptide complexes as the infection progresses into chronicity.

The NetMHCpan4.0 algorithm, which is based on binding affinity and integrates data on eluted naturally processed ligands, reflected optimal HLA class I binding for 9-mers, producing a decreasing number of predictions when the peptide size was increased from 9 to 11 amino acids, with minor differences for further increases between sizes of 11, 12, 13 and 14 amino acids. With a single exception, predicted binders between 11-14 amino acids included at least one 9mer predicted to bind on its own, suggesting that a destabilizing effect of the extra amino acids beyond the canonical HLA class I binding pockets at positions 2 and 9 could account for fewer predictions.

Important limitations of this study are the mismatch between sizes and scanning frames of previously experimentally evaluated (17mers overlapping by 11 aa) and algorithmically

predicted peptides (9-14 mers overlapping by 8-14aa), and lack of predictions of HLA class II restricted epitopes, which might have contributed to a fraction of IFN-g ELISPOT responses.
These could be addressed by further developments of the prediction algorithm.

Conclusions

In this analysis, using NetMHCpan4.0 to predict previously experimentally mapped epitopes, we demonstrate that the computational method is reliable and predicts an acceptable portion of binder epitopes. We recommend the use of such computational methods to reduce the size of experiments required therefore markedly reducing the time and cost associated.

Methods

Experimental Binder Data

Experimental data of peptides previously mapped for HIV-1 epitope recognition of 6 individuals (Table 4) at 2 time points each was used for comparison with the positive computationally predicted binders. These were from a Ugandan early HIV-1 serodiscordant couple cohort approved by the Uganda Virus Research Institute (UVRI) Research and Ethics review board and the Uganda National Council of Science and Technology (UNCST). All participants provided informed consent. Six (6) participants whose experimental epitope recognition profile we evaluated were early HIV-1 infections (Table 4), enrolled under the following criteria: (i) detection of HIV-1 P24 antigen with a simultaneous negative HIV-1 antibody ELISA (2 participants) or documented HIV-1 sero-negative test in the previous 12 months (4 participants); (ii) HAART naïve (all). Early infection was determined following the Fiebig Staging criteria[21] as described elsewhere by Obuku A.E. *et.al.*[22].
The experimentally tested peptides totalled 757 (figure 5), were 17aa long, overlapping by 11aa and spanning the HIV-1 proteome consensus for subtypes A and D. Cultured ELISPOT assays using 200000 cells/well as previously documented by Obuku AE. *et.al.* [22] and *ex-vivo* IFN-g ELISPOT assay using 100000 cells/well were used for testing peptide pools and epitope mapping respectively. Experimental positive pools were 3 times the background

wells and at least 600 spot forming units per million cells. “Deconvolute This” software [23] was used to identify possible responding individual peptides from the pools or where it was not possible all the peptides in a pool were tested as single peptides (figure 6).

HLA typing

High resolution reference strand conformation analysis HLA class I tissue typing for the early infected subjects was done using methods described elsewhere [24].

HIV-1 subtyping

HIV-1 subtyping determination was performed on the *gag* gene [25, 26] using Sanger method generated sequences. The sequences were input into the REGA HIV-1 automated subtyping tool to determine the HIV-1 clade [27, 28].

Computation Epitope Prediction.

The HIV-1 A and D subtype consensus sequences were used as inputs for the computational epitope prediction. The web version of NetMHCpan4.0 [12] (http://www.cbs.dtu.dk/services/NetMHCpan/) was configured to predict 9mer through 14mer epitopes for 20 HLA class I alleles (Table 4) that were expressed by the 6 HIV infected donors. Perl version 5.26.2 was used to extract the binders from all the NetMHCpan4.0 predictions and also to compare the computational binders to the 93 mapped
experimental 17aa peptides for 9mer through 14mer hits using a sliding window. An experimental peptide was considered a hit if any of the computational 9mer through 14mer sequence was contained in the 17 amino acid experimental peptide sequence as well as any of the HLA-A, B or C expressed by the individual matched the NetMHCpan4.0 HLA class I type(s). If multiple computational epitope predictions were contained in a single 17mer experimental peptide they counted as a single hit. These were determined by a BLAST search of the computational binders against the derivative experimental peptides to determine computational predictions from the same test peptide.

Data Analysis

Statistics computations and plots were generated using SPSS version 24.0.0.0. The NetMHCpan4.0 computational performance was evaluated using a confusion matrix to classify true positives, true negatives, false positives and false negatives that were used for the Receiver Operator Characteristic (ROC) plot. The hit rate (sensitivity) and false hit rate (specificity) of binder predictions as determined by the NetMHCpan4.0 threshold of peptides within the top 2% (with a score of 2 or less) were calculated and the strength of the model was determined by calculating the area under the curve, AUC of the ROC plot [29-31]. Pearson's correlation coefficient was used to evaluate the relationship between the number of epitopes with various HIV-1 genes. To evaluate if there were any differences in the early versus late time point peptides for the binding ranking of the experimentally mapped peptides as predicted by the computational score the Wilcoxon signed rank test was used. To evaluate if HIV-1 subtypes A and D affected the number of computational predictions generated, Fisher Exact Test was used. To determine whether multiple computationally predicted epitope sequences were derived from the same experimental peptide sequence, a local blast database was set up using Geneious version 9.0.5. Both clades A and D experimental consensus sequences were used separately each as a reference sequence for the blast. The computational peptide sequences were then aligned against the consensuses to evaluate those derived from a single 17 amino acid experimental peptide sequence. Where an experimental peptide was predicted by multiple or overlapping computational peptides, the average NetMHCpan4.0 score was assigned as the computational score for this peptide. This score was also used during the generation of the ROC curve and the confusion matrix.

Abbreviations

AUC – Area under the curve

CTL – Cytotoxic T lymphocytes

HIV-1 – Human immunodeficiency virus type 1

HLA – Human leucocyte antigen

ROC – Receiver operator characteristic

Declarations

Acknowledgement.

We thank the study participants who provided specimen for the wet laboratory experiments, the AIDS Information Centre Clinic Kampala, Uganda that steered the Rubicon discordant couple cohort study and the late Anthony Kebba who initiated the Rubicon Cohort.

We thank Ruhena Sargeant for HLA typing, the late Harr F. Njai for HIV-1 ELISA and Western blot assays and Deogratius Ssemwanga for help with GenBank submissions.

Funding.

This research is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement, the Wellcome Trust (grant WT078927MA), and EDCTP (project code: TA_05_40200_40203).

Author's contributions.

DBL performed the peptide mapping experiments. CW recruited participants. NN provided HIV subtyping. DBL analysed the data together with PP. DBL wrote the manuscript with contributions from PP, AEO and JS. All authors reviewed the manuscript and or provided useful contributions as well as approved the final manuscript.

Ethics approval and consent to participate.

The Rubicon study, from which we derived the experimental data, was reviewed and approved by the Uganda Virus Research Institute, Research and Ethics Committee (UVRI-REC) and the Uganda National Council of Science and Technology (UNCT). All participants provided informed signed consent accepting to freely participate in this study.

Competing interests.

The authors declare that they have no competing interests.

References

1. Tambunan US, Sipahutar FR, Parikesit AA, Kerami D: **Vaccine Design for H5N1 Based on B- and T-cell Epitope Predictions**. *Bioinformatics and biology insights* 2016, **10**:27-35.
2. Yu K, Petrovsky N, Schonbach C, Koh JY, Brusic V: **Methods for prediction of peptide binding to MHC molecules: a comparative study**. *Molecular medicine* 2002, **8**(3):137-148.
3. Azizi A, Anderson DE, Torres JV, Ogrel A, Ghorbani M, Soare C, Sandstrom P, Fournier J, Diaz-Mitoma F: **Induction of broad cross-subtype-specific HIV-1 immune responses by a novel multivalent HIV-1 peptide vaccine in cynomolgus macaques**. *Journal of immunology* 2008, **180**(4):2174-2186.
4. Baden LR, Walsh SR, Seaman MS, Cohen YZ, Johnson JA, Licona JH, Filter RD, Kleinjan JA, Gothing JA, Jennings J *et al*: **First-in-Human Randomized Controlled Trial of Mosaic HIV-1 Immunogens Delivered via a Modified Vaccinia Ankara Vector**. *J Infect Dis* 2018.
5. Yebra G, Ragonnet-Cronin M, Ssemwanga D, Parry CM, Logue CH, Cane PA, Kaleebu P, Brown AJ: **Analysis of the history and spread of HIV-1 in Uganda using phylodynamics**. *J Gen Virol* 2015, **96**(Pt 7):1890-1898.
6. Lim WC, Khan AM: **Mapping HLA-A2, -A3 and -B7 supertype-restricted T-cell epitopes in the ebolavirus proteome**. *BMC genomics* 2018, **19**(Suppl 1):42.
7. Goodswen SJ, Kennedy PJ, Ellis JT: **Enhancing in silico protein-based vaccine discovery for eukaryotic pathogens using predicted peptide-MHC binding and peptide conservation scores**. *PloS one* 2014, **9**(12):e115745.
8. Goodswen SJ, Kennedy PJ, Ellis JT: **Discovering a vaccine against neosporosis using computers: is it feasible?** *Trends in parasitology* 2014, **30**(8):401-411.
9. Goodswen SJ, Kennedy PJ, Ellis JT: **Vacceed: a high-throughput in silico vaccine candidate discovery pipeline for eukaryotic pathogens based on reverse vaccinology**. *Bioinformatics* 2014, **30**(16):2381-2383.
10. Pedersen LE, Rasmussen M, Harndahl M, Nielsen M, Buus S, Jungersen G: **A combined prediction strategy increases identification of peptides bound with high affinity and stability to porcine MHC class I molecules SLA-1*04:01, SLA-2*04:01, and SLA-3*04:01**. *Immunogenetics* 2016, **68**(2):157-165.
11. 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.
12. 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**. *Journal of immunology* 2017, **199**(9):3360-3368.
13. Zhang H, Lundegaard C, Nielsen M: **Pan-specific MHC class I predictors: a benchmark of HLA class I pan-specific prediction methods**. *Bioinformatics* 2009, **25**(1):83-89.
14. Peters B, Bui HH, Frankild S, Nielson M, Lundegaard C, Kostem E, Basch D, Lamberth K, Harndahl M, Fleri W *et al*: **A community resource benchmarking predictions of peptide binding to MHC-I molecules**. *PLoS computational biology* 2006, **2**(6):e65.

15. Jacob L, Vert JP: **Efficient peptide-MHC-I binding prediction for alleles with few known binders**. *Bioinformatics* 2008, **24**(3):358-366.
16. Zhang GL, Khan AM, Srinivasan KN, August JT, Brusic V: **MULTIPRED: a computational system for prediction of promiscuous HLA binding peptides**. *Nucleic acids research* 2005, **33**(Web Server issue):W172-179.
17. DeLisi C, Berzofsky JA: **T-cell antigenic sites tend to be amphipathic structures**. *Proceedings of the National Academy of Sciences of the United States of America* 1985, **82**(20):7048-7052.
18. He Y, Rappuoli R, De Groot AS, Chen RT: **Emerging vaccine informatics**. *Journal of biomedicine & biotechnology* 2010, **2010**:218590.
19. Reche PA, Reinherz EL: **Prediction of peptide-MHC binding using profiles**. *Methods in molecular biology* 2007, **409**:185-200.
20. Roider J, Meissner T, Kraut F, Vollbrecht T, Stirner R, Bogner JR, Draenert R: **Comparison of experimental fine-mapping to in silico prediction results of HIV-1 epitopes reveals ongoing need for mapping experiments**. *Immunology* 2014, **143**(2):193-201.
21. Fiebig EW, Wright DJ, Rawal BD, Garrett PE, Schumacher RT, Peddada L, Heldebrant C, Smith R, Conrad A, Kleinman SH *et al*: **Dynamics of HIV viremia and antibody seroconversion in plasma donors: implications for diagnosis and staging of primary HIV infection**. *Aids* 2003, **17**(13):1871-1879.
22. Obuku AE, Bugembe DL, Musinguzi K, Watera C, Serwanga J, Ndembi N, Levin J, Kaleebu P, Pala P: **Macrophage Inflammatory Protein-1 Beta and Interferon Gamma Responses in Ugandans with HIV-1 Acute/Early Infections**. *AIDS research and human retroviruses* 2016, **32**(3):237-246.
23. Roederer M, Koup RA: **Optimized determination of T cell epitope responses**. *Journal of immunological methods* 2003, **274**(1-2):221-228.
24. Arguello JR, Little AM, Bohan E, Goldman JM, Marsh SG, Madrigal JA: **High resolution HLA class I typing by reference strand mediated conformation analysis (RSCA)**. *Tissue antigens* 1998, **52**(1):57-66.
25. Serwanga J, Mugaba S, Pimego E, Nanteza B, Lyagoba F, Nakubulwa S, Heath L, Nsubuga RN, Ndembi N, Gotch F *et al*: **Profile of T cell recognition of HIV type 1 consensus group M Gag and Nef peptides in a clade A1- and D-infected Ugandan population**. *AIDS research and human retroviruses* 2012, **28**(4):384-392.
26. Serwanga J, Nakiboneka R, Mugaba S, Magambo B, Ndembi N, Gotch F, Kaleebu P: **Frequencies of Gag-restricted T-cell escape "footprints" differ across HIV-1 clades A1 and D chronically infected Ugandans irrespective of host HLA B alleles**. *Vaccine* 2015, **33**(14):1664-1672.
27. Alcantara LC, Cassol S, Libin P, Deforche K, Pybus OG, Van Ranst M, Galvao-Castro B, Vandamme AM, de Oliveira T: **A standardized framework for accurate, high-throughput genotyping of recombinant and non-recombinant viral sequences**. *Nucleic acids research* 2009, **37**(Web Server issue):W634-642.

28. de Oliveira T, Deforche K, Cassol S, Salminen M, Paraskevis D, Seebregts C, Snoeck J, van Rensburg EJ, Wensing AM, van de Vijver DA *et al*: **An automated genotyping system for analysis of HIV-1 and other microbial sequences**. *Bioinformatics* 2005, **21**(19):3797-3800.
29. Hanley JA, McNeil BJ: **The meaning and use of the area under a receiver operating characteristic (ROC) curve**. *Radiology* 1982, **143**(1):29-36.
30. Park SH, Goo JM, Jo CH: **Receiver operating characteristic (ROC) curve: practical review for radiologists**. *Korean journal of radiology* 2004, **5**(1):11-18.
31. Hanczar B, Hua J, Sima C, Weinstein J, Bittner M, Dougherty ER: **Small-sample precision of ROC-related estimates**. *Bioinformatics* 2010, **26**(6):822-830.

Tables

Table 1. Experimental and computational 9mer peptide confusion matrix.

The total number of peptides experimentally tested were 757 and these are broken down to show the fractions from both the experimental testing and NetMHCpan4.0 computational predictions.

Table 2. Experimentally Mapped Peptides and Computationally Predicted Epitopes.

ID	Participant's HLA Types	Hits	Screening Peptide	NetMHCpan4.0 9-mer Epitope Prediction	NetMHCpan4.0 9-mer HLA Prediction	NetMHCpan4.0 % Rank
E91	A*02:01	1	FITKGLGISYGRKKRRQR	GLGISYGRK	A*0301	0.5
	A*03:01	2	HPKVSSEVHIPLGDARLV	IPLGDARLV	B*53:01	0.7
	B*53:01		HPKVSSEVHIPLGDARLV	KVSSEVHIP	B*58:02	0.6
	B*58:02		HPKVSSEVHIPLGDARLV	SSEVHIPLG	B*58:02	0.6
	Cw*04:01		HPKVSSEVHIPLGDARLV	VSSEVHIPL	Cw*04:01	0.9
	Cw*06:02		HPKVSSEVHIPLGDARLV	HPKVSSEVH	B*53:01	1.2
E92	A*02:01	3	RKQNPEIVIYQYMDDLYV	YQYMDDLYV	A*02:02	0.15
	A*30:02		RKQNPEIVIYQYMDDLYV	NPEIVIYQY	B*44:03	0.6
	B*44:03		RKQNPEIVIYQYMDDLYV	YQYMDDLYV	A*02:01	1.8
	B*14:02		RKQNPEIVIYQYMDDLYV	YQYMDDLYV	Cw*04:01	0.9
	Cw*04:01		RKQNPEIVIYQYMDDLYV	YQYMDDLYV	Cw*08:02	1.2
	Cw*08:02		RKQNPEIVIYQYMDDLYV	VIYQYMDDL	A*30:02	0.6
		4	ELNKRTQDFWEVQLGIPH	TQDFWEVQL	Cw*08:02	0.4
			ELNKRTQDFWEVQLGIPH	TQDFWEVQL	A*02:01	0.03
			ELNKRTQDFWEVQLGIPH	TQDFWEVQL	Cw*08:02	1.5
			ELNKRTQDFWEVQLGIPH	ELNKRTQDF	B*44:03	0.6
		5	NDIQKLVGKLNWASQIYP	KLNWASQIY	A*30:02	0.5
			NDIQKLVGKLNWASQIYP	KLVGKLNWA	A*02:01	0.47
			NDIQKLVGKLNWASQIYP	KLNWASQIY	Cw*04:01	1.8
		6	PAIQTGSEELRSLYNTVA	GSEELRSLY	A*30:02	0.4
			PAIQTGSEELRSLYNTVA	SEELRSLY	B*44:03	0.5
		7	PYNTPIFAIKKKDSTKWR	PYNTPIFAI	Cw*04:01	1.14
		8	GANNSHNETFRPGGGDMR	TFRPGGGDM	Cw*04:01	1.6
		9	GMDGPKVKQWPLTEEKIK	MDGPKVKQW	B*44:03	0.5
		10	PLTSLKSLFGNDPLSQ	KSLFGNDPL	A*02:01	1.4
			PLTSLKSLFGNDPLSQ	KSLFGNDPL	Cw*08:02	1.2
			FKGPRKIIKCFNCGKEGHI			
		11	HERIEVKDTKEALEKI	EVKDTKEAL	B*14:02	1.4
			HERIEVKDTKEALEKI	IEVKDTKEA	B*44:03	1.99
E94	A*34:02	12	KIEEIQNKSKQKTQQAAA	EIQNKSKQK	A*34:02	1.03
	A*74:01	13	NHPSCVWLEAQEEEEVGF	LEAQEEEEV	B*44:15	1.7
	B*44:03	14	HQDPIPKQPSSQPRGD	HQDPIPKQP	Cw*04:01	0.6
	B*58:02					
	Cw*04:01					
	Cw*06:02					
E95	A*23:01	15	VAVHVASGYIEAEVIPA	VAVHVASGY	Cw*16:01	1.5
	A*74:01	17	KRWIILGLNKIVRMYSPV	WIILGLNKI	A*23:01	0.6
	B*44:03	16	KRWIILGLNKIVRMYSPV	IILGLNKIV	A*74:01	0.6
	B*15:10	17	NMMLNIVGGHQAAMQMLK	HQAAMQMLK	B*15:10	0.17
	Cw*04:01		NMMLNIVGGHQAAMQMLK	HQAAMQMLK	A*74:01	0.9
	Cw*16:01		NMMLNIVGGHQAAMQMLK	HQAAMQMLK	Cw*04:01	1.3
		18	KNWMTETLLVQNANPDCK	TETLLVQNA	B*44:15	0.09
			KNWMTETLLVQNANPDCK	KNWMTETLL	A*23:01	0.8
		19	FRDYVDRFFKTLRAEQA	FRDYVDRFF	Cw*04:01	0.03

			FRDYVDRFFKTLRAEQA	FRDYVDRFF	A*23:01	0.6
			FRDYVDRFFKTLRAEQA	FRDYVDRFF	Cw*04:01	1.1
		20	GATL**EEMMTACQG**VGGPGH	EEMMTACQG	B*44:03	0.25
		21	L**RALGPGATL**EEMMTA	RALGPGATL	B*15:10	1.8
			L**RALGPGATL**EEMMTA	RALGPGATL	Cw*04:01	0.6
		22	FFKTLR**AEQATQEV**KNWM	AEQATQEVK	B*44:03	0.15
		23	MEKEGKI**SKIGPENPY**	SKIGPENPY	B*15:03	0.5
			MEKEGKI**SKIGPENPY**	SKIGPENPY	B*15:10	0.5
		24	WVKVIEEK**AFSPEVIPMF**	AFSPEVIPMF	A*23:01	0.4
			WVKVIEEKAFSPEVIPMF	WVKVIEEKA	A*23:01	1.7
			WVKVI**EEKAFSPEV**IPMF	EEKAFSPEV	B*44:03	0.8
			WVKVI**EEKAFSPEV**IPMF	EEKAFSPEV	B*44:15	0.03
			WVKVIEEKA**FSPEVIPMF**	FSPEVIPMF	Cw*04:01	0.5
			WVKVIEEKA**FSPEVIPMF**	FSPEVIPMF	Cw*16:01	0.5
			WVKVIEE**KAFSPEVIP**MF	KAFSPEVIP	Cw*16:01	1.2
		25	HQMKDCTE**RQANFLGKI**W	RQANFLGKI	B*4403	1
		26	PMFSAL**SEGATPQDL**NMM	SEGATPQDL	B*44:03	0.8
		27	**HLARNCRAP**RKKGCWK	HLARNCRAP	A*74:01	0.6
			LVQNANPDCKSILRAL			
		28	V**ATLYCVHQR**IDVKDTK	ATLYCVHQR	A*74:01	0.9
		29	KIE**EIQNKSKQK**TQQAAA	EIQNKSKQK	A*74:01	1.03
		30	**AGPIPPGQM**REPRGSDIA	AGPIPPGQM	B*15:10	0.6
			SKQKTQQAAADTGNSSKV			
E913	A*02:01	31	**LWQRPLVTI**KIGGQLKEA	LWQRPLVTI	A*02:01	1.6
	A*34:02		LW**QRPLVTIK**IGGQLKEA	QRPLVTIKI	Cw*06:02	0.7
	B*45:01		L**WQRPLVTIK**IGGQLKEA	WQRPLVTIK	B*47:01	1.9
	B*47:01	32	TVPVK**LKPGMDGPK**VKQW	LKPGMDGPK	A*34:02	0.9
	Cw*06:02					
	Cw*16:01					
E914	A*01:01	33	DKWASLWNW**FSITQWLW**Y	FSITQWLWY	A*01:01	0.06
	A*02:01		D**KWASLWNWF**SITQWLWY	KWASLWNWF	Cw*04:07	1.2
	B*07:02		DKWA**SLWNWFSIT**QWLWY	SLWNWFSIT	A*02:01	1.66
	B*44:15	34	PVDPDEVEK**ATEGENNSL**	ATEGENNSL	A*01:01	1.74
	Cw*04:07					
	Cw*07:02					
L91	A*02:01	35	EQMHTD**IISLWDQSLK**	IISLWDQSLK	A*03:01	1.9
	A*03:01		EQ**MHTDIISLW**DQSLK	MHTDIISLW	B*58:02	0.9
	B*53:01		EQ**MHTDIISLW**DQSLK	MHTDIISLW	Cw*06:02	1.3
	B*58:02		EQ**MHTDIISL**WDQSLK	QMHTDIISL	A*02:01	1.9
	Cw*04:01		EQ**MHTDIISL**WDQSLK	QMHTDIISL	B*5301	1.2
	Cw*06:02	36	LETSEGCKQ**IIGQLQPAI**	IIGQLQPAI	A*02:01	0.4
		37	SGGKLDAWE**KIRLRPGGK**	KIRLRPGGK	A*03:01	0.25
		38	LETTEGCQQ**IMEQLQPAL**	IMEQLQPAL	A*03:01	0.4
			LETTEGCQQ**IMEQLQPAL**	IMEQLQPAL	Cw*04:01	0.8
			LETTEGCQ**QIMEQLQPA**L	QIMEQLQPA	A*02:01	0.7
		39	ERILSTCLG**RSAEPVPL**	RSAEPVPL	B*58:02	0.12
			E**RILSTCLGR**SAEPVPL	RILSTCLGR	A*03:01	0.9

			ERILSTCLGRSAEPVPL	CLGRSAEPV	A*02:01	2
		40	LVGPTPVNIIGRNMLTQI	LVGPTPVNI	A*02:01	1.61
		41	CKQIIGQLQPAIQTGSEEL	QIIGQLQPA	A*02:01	1.8
			CKQIIGQLQPAIQTGSEEL	AIQTGSEEL	A*03:01	1.5
			CKQIIGQLQPAIQTGSEEL	IIGQLQPAI	A*03:01	1.1
		42	PAIQTGSEELRSLYNTVA	AIQTGSEEL	A*03:01	1.5
			PAIQTGSEELRSLYNTVA	LRSLYNTVA	Cw*06:02	0.7
L92	A*02:01	43	NDIQKLVGKLNWASQIYP	KLNWASQIY	A*30:02	0.5
	A*30:02		NDIQKLVGKLNWASQIYP	KLVGKLNWA	A*02:01	0.6
	B*44:03		NDIQKLVGKLNWASQIYP	KLNWASQIY	A*02:01	0.9
	B*14:02		NDIQKLVGKLNWASQIYP	KLNWASQIY	Cw*04:01	1.8
	Cw*04:01	44	LVVKTYWGLHTGEREWHL	LVVKTYWGL	A*02:01	1.7
	Cw*08:02		LVVKTYWGLHTGEREWHL	VVKTYWGLH	A*30:02	1.5
		45	SLVNRVRQGYSPLSFQTL	NRVRQGYSPL	B*14:02	0.12
			SLVNRVRQGYSPLSFQTL	YSPLSFQTL	Cw*04:01	0.7
			SLVNRVRQGYSPLSFQTL	RQGYSPLSF	A*30:02	1.2
			SLVNRVRQGYSPLSFQTL	RQGYSPLSF	Cw*04:01	1.4
		46	TLPCRIKQIINMWQGV	CRIKQIINM	A*02:02	0.4
		47	MRVRGIQRNYQHLWRW	RNYQHLWRW	B*44:03	0.4
		48	GEMKNCSFNITTEIRDKK	EMKNCSFNI	B*44:03	0.3
		49	NVTENFNMWKNNMVEQMH	NFNMWKNNM	Cw*04:01	1.06
			NVTENFNMWKNNMVEQMH	TENFNMWKNNM	B*44:03	1.81
		50	WLIDRIRERAEDSGNESE	WLIDRIRER	A*02:01	2
L94	A*3402	51	LIHLHYFDCFSDSAIRKA	YFDCFSDSA	Cw*04:01	0.9
	A*7401		LIHLHYFDCFSDSAIRKA	YFDCFSDSA	Cw*06:02	1.6
	B*4403		LIHLHYFDCFSDSAIRKA	HLHYFDCFSDSAIR	A*7401	1.4
	B*5802		LIHLHYFDCFSDSAIRKA	FSDSAIRKA	Cw*04:01	1.1
	Cw*0401	52	HLARNCRAPRKKGCWK	ARNCRAPRK	A*3402	1.1
	Cw*0602		HLARNCRAPRKKGCWK	HLARNCRAP	A*3402	1.5
			HLARNCRAPRKKGCWK	HLARNCRAP	A*74:01	0.6
		53	SKQKTQQAAADTGNSSKV	AADTGNSSK	A*3402	1.15
		54	HQDPIPKQPSSQPRGD	HQDPIPKQP	Cw*04:01	0.6
L95	A*23:01	55	KRWIILGLNKIVRMYSPV	WIILGLNKI	A*23:01	0.6
	A*74:01		KRWIILGLNKIVRMYSPV	IILGLNKIV	A*74:01	0.6
	B*44:03	56	NMMLNIVGGHQAAMQMLK	GHQAAMQML	B*15:10	0.4
	B*15:10		NMMLNIVGGHQAAMQMLK	HQAAMQMLK	A*74:01	0.9
	Cw*04:01		NMMLNIVGGHQAAMQMLK	HQAAMQMLK	Cw*04:01	1.3
	Cw*16:01	57	EVNIVTDSQYALGIIQA	EVNIVTDSQ	B*44:03	0.5
		58	AYETEMHNVWATHACV	TEMHNVWAT	B*44:03	0.4
			AYETEMHNVWATHACV	MHNVWATHA	B*15:10	0.9
			AYETEMHNVWATHACV	MHNVWATHA	B*44:03	0.9
		59	AAEWDRLHPVHAGPI	AAEWDRLHP	B*44:03	0.5
			AAEWDRLHPVHAGPI	LHPVHAGPI	B*15:10	0.15
			AAEWDRLHPVHAGPI	RLHPVHAGP	A*74:01	1.5
		60	LRALGPGATLEEMMTA	RALGPGATL	B*15:10	0.6
			LRALGPGATLEEMMTA	RALGPGATL	Cw*04:01	0.6
		61	FFKTLRAEQATQEVKNWM	AEQATQEVK	B*44:03	0.15

		62	GTTSTP**QEQIGWMTG**NPPI	QEQIGWMTG	B*44:15	0.65
			GTTSTPQEQI**GWMTGNPPI**	GWMTGNPPI	A*23:01	1.4
		63	WVKVI**EEKAFSPEV**IPMF	EEKAFSPEV	B*44:15	0.62
			WVKVI**EEKAFSPEV**IPMF	EEKAFSPEV	B*44:03	0.8
			WVKVIEEKAFSPEVIPMF	WVKVIEEKA	A*23:01	1.7
			WVKVIEEKA**FSPEVIPMF**	FSPEVIPMF	Cw*04:01	0.5
			WVKVIEEK**AFSPEVIPMF**	AFSPEVIPM	Cw*16:01	0.7
			WVKVIEE**KAFSPEVIP**MF	KAFSPEVIP	Cw*16:01	1.2
		64	**TVYYGVPVW**KDAETTLF	TVYYGVPVW	A*74:01	0.17
			TVYYGVPVWKDAETTLF	TVYYGVPVW	Cw*16:01	1.1
			TVYYGVPV**WKDAETTLF**	WKDAETTLF	B*15:10	0.9
			TVYYGVP**VWKDAETTL**F	VWKDAETTL	A*23:01	0.6
			TVYYGVP**VWKDAETTL**F	VWKDAETTL	Cw*04:01	0.6
		65	L**RWGTMILGM**IIICSAA	RWGTMILGM	A*23:01	0.9
			L**RWGTMILGM**IIICSAA	RWGTMILGM	Cw*04:01	0.94
		66	G**HQAAMQMLK**DTINEEAA	HQAAMQMLK	A*74:01	0.9
		67	IKQGPKEP**FRDYVDRFF**K	FRDYVDRFF	A*23:01	0.6
			IKQGPKEP**FRDYVDRFF**K	FRDYVDRFF	Cw*04:01	0.6
			IKQGPKEP**FRDYVDRFF**K	FRDYVDRFF	Cw*04:01	1.1
		68	**FRDYVDRFF**KTLRAEQA	FRDYVDRFF	A*23:01	0.6
			LVQNANPDCKSILRAL			
		69	**MREPRGSDI**AGTTSTPQEQI	MREPRGSDI	B*15:10	2
		70	EKI**RLRPGGKKK**YRLKHL	RLRPGGKKK	A*74:01	0.28
		71	V**ATLYCVHQR**IDVKDTK	ATLYCVHQR	A*74:01	0.9
		72	LFCA**SDAKAYETEMHNVW**	SDAKAYETEMHNVW	B*44:03	0.31
L913	A*02:01	73	PP**LVKLWYQLE**KEPIIGA	LVKLWYQLE	A*34:01	0.5
	A*34:02		PPLVKLWY**QLEKEPIIG**A	QLEKEPIIG	B*45:01	0.9
	B*45:01		PPLVKLW**YQLEKEPII**GA	YQLEKEPII	A*02:01	0.8
	B*47:01		PPLVKLW**YQLEKEPII**GA	YQLEKEPII	B*47:01	1.2
	Cw*06:02	74	KWKPK**MIGGIGGFIK**VR	MIGGIGGFIK	A*34:02	0.2
	Cw*16:01		KWKP**KMIGGIGGF**IKVR	KMIGGIGGF	A*02:01	1
			KWKP**KMIGGIGGF**IKVR	KMIGGIGGF	B*47:01	1.1
		75	**VIWGKTPK**FRLPIQKETW	IVIWGKTPK	A*34:02	0.15
			VIWG**KTPKFRLPI**QKETW	KTPKFRLPI	Cw*16:01	1.1
			VIWGKTPKFRLPIQKETW	VIWGKTPKF	A*34:02	1
		76	**RQANFLGKI**WPSHKGR	RQANFLGKI	B*47:01	0.4
			RQANFLGKIWPSHKGR	RQANFLGKI	Cw*06:02	2
			RQAN**FLGKIWPSH**KGR	FLGKIWPSH	A*34:02	1.1
		77	KIEEL**REHLLRWGF**TTPDK	REHLLRWGF	B*47:01	0.03
			KIEEL**REHLLRWGF**TTPDK	REHLLRWGF	B*45:01	0.7
			KIEE**LREHLLRWGF**TTPDK	LREHLLRWG	Cw*06:02	1.4
			KIEELRE**HLLRWGFTT**PDK	HLLRWGFTT	A*02:01	1.2
		78	GFAILKCKD**KEFNGTGPC**K	KEFNGTGPC	B*45:01	1.5
		79	AILNIPTR**IRQGLERA**LL	IRQGLERAL	Cw*06:02	0.6
			AILNIPTRIRQGLERALL	AILNIPTRI	A*02:01	0.6
			AILNIPTRI**RQGLERALL**	RQGLERALL	B*47:01	1.7
		80	QKTELQAIN**LALQDSGLE**V	LALQDSGLE	A*02:01	1.5

			QKTELQAINLALQDSGLEV	TELQAINLA	B*47:01	0.6
			QKTELQAINLALQDSGLEV	QKTELQAIN	B*45:01	1
			QKTELQAINLALQDSGLEV	NLALQDSGL	A*34:02	1.5
		81	IIGRNLLTQIGCTLNFPI	IGCTLNFPI	A*02:01	0.9
			IIGRNLLTQIGCTLNFPI	LLTQIGCTL	A*02:01	1.9
			IIGRNLLTQIGCTLNFPI	TQIGCTLNF	Cw*16:01	1.4
			IIGRNLLTQIGCTLNFPI	LLTQIGCTL	Cw*16:01	0.9
		82	KWKPKMIGGIGGFIKVR	KMIGGIGGF	A*02:01	1
		83	LWQRPLVTIKIGGQLKEA	LWQRPLVTI	A*02:01	1.6
			LWQRPLVTIKIGGQLKEA	QRPLVTIKI	Cw*06:02	0.7
		84	LKEALLDTGADDTVLEEI	LKEALLDTG	B*45:01	1.2
		85	KRQEILDLWVYHTQGYF	QEILDLWVY	B*45:01	1.7
			KRQEILDLWVYHTQGYF	QEILDLWVY	B*47:01	0.9
			KRQEILDLWVYHTQGYF	RQEILDLWV	Cw*06:02	1.1
			KRQEILDLWVYHTQGYF	ILDLWVYHT	A*02:01	0.7
			IYSLIEESQNQQEKNEQEL			
L914	A*01:01	86	SFNCGGEFFYCNTSGLF	SFNCGGEFFY	A*01:01	0.25
	A*02:01		SFNCGGEFFYCNTSGLF	SFNCGGEFF	Cw*04:07	1.3
	B*07:02		SFNCGGEFFYCNTSGLF	SFNCGGEFF	Cw*07:02	1.1
	B*44:03		SFNCGGEFFYCNTSGLF	GEFFYCNTS	B*44:03	1.3
	Cw*04:07	87	MEKEGKISKIGPENPY	KEGKISKIGPENPY	B*44:03	1.3
	Cw*07:02		MEKEGKISKIGPENPY	ISKIGPENP	A*01:01	1.8
		88	ARKNRRRRWRARQRQI	RRWRARQRQ	Cw*07:02	0.6

Experimentally mapped peptides for all participants and their cognate computational core 9-mer and a single 14-mer epitope sequence with scores. Peptides shown in *italic* text were not algorithmically predicted as binders. Multiple computational predictions contained in a single experimental peptide were counted as a single hit. Participant's identifiers (ID) beginning with E or L represent early or late time sampling points respectively.

Table 3. Peptides not predicted

Participant's Identification	Participant's HLA Alleles	Experimental Peptide Sequence
E92	A*02:01	FKGPRKIIKCFNCGKEGHI
	A*30:02	
	B*44:03	
	B*14:02	
	Cw*04:01	
	Cw*08:02	
E95	A*23:01	LVQNANPDCKSILRAL (both time points)
	A*74:01	SKQKTQQAAADTGNSSKV
	B*44:03	
	B*15:10	
	Cw*04:01	
	Cw*16:01	
L913	A*02:01	IYSLIEESQNQQEKNEQEL
	A*34:02	
	B*45:01	
	B*47:01	
	Cw*06:02	
	Cw*16:01	

Experimentally mapped peptides that were not predicted by NetMHCpan4.0 binders. Participant's identifiers beginning with E or L represent early or late time sampling points respectively.

Table 4. Participant characteristics, HIV-1 infecting stage clade and HLA class I haplotypes.

Subject	Sex	Age range (years)	HIV-1 subtype	Class-I HLA	Early Time Point (Days)	Fiebig Staging	Late Time Point (Days)
91	M	31-40	A	A*0201,*0301;B*5301,*5802;Cw*0401,*0602	121	VI	841
92	F	21-30	D	A*0201,*3002;B*4403,*1402Cw*0401,*0802	52	VI	743
94	M	51-60	A	A*3402,*7401;B*4403,*5802;Cw*0401,*0602	28	V	358
95	M	21-30	A	A*2301,*7401;B*4403,*1510;Cw*0401,*1601	30	VI	570
913	F	11-20	D	A*0201,*3402;B*4501,*4701;Cw*0602,*1601	61	VI	211
914	F	21-30	D	A*0101,*0201;B*0702,*4415;Cw*0407,*0702	31	IV	181

Figures

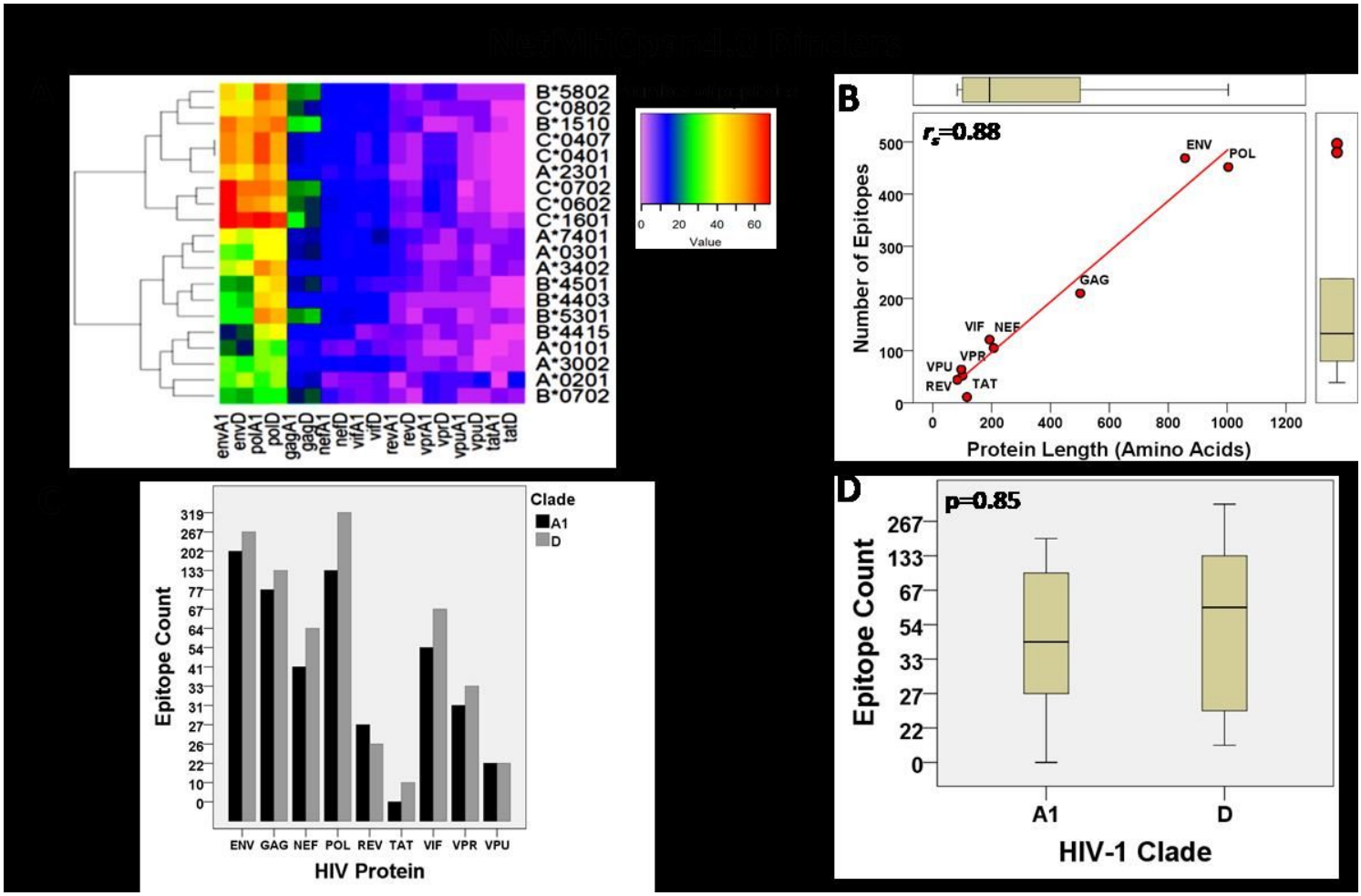

Figure 1

NetMHCpan Binder Predictions. (A) Heatmap showing absolute count of NetMHCpan4.0 predicted 9-mer binders for all HLA types against HIV-1 genes. The dendrogram shows the nearest similarity for the number of predicted counts across HLA types; (B) the length of the HIV-1 protein sequence plotted against the absolute number of NetMHCpan4.0 predicted 9mer binders showing a positive correlation (Spearman's correlation coefficient, rs=0.88). The number of computational predictions for HIV-1 is dependent on the length of the sequence; (C) comparison of HIV-1 clade A and D absolute number of NetMHCpan4.0 predicted 9mer binders per HIV-1 gene for the wet experiment test peptide sequences. The algorithm predicted more binders for Clade D than clade A; (D) the difference in the number of NetMHCpan4.0 predictions between HIV-1 clade A and D did not approach statistical significance (Fishers Exact Test; p=0.085).

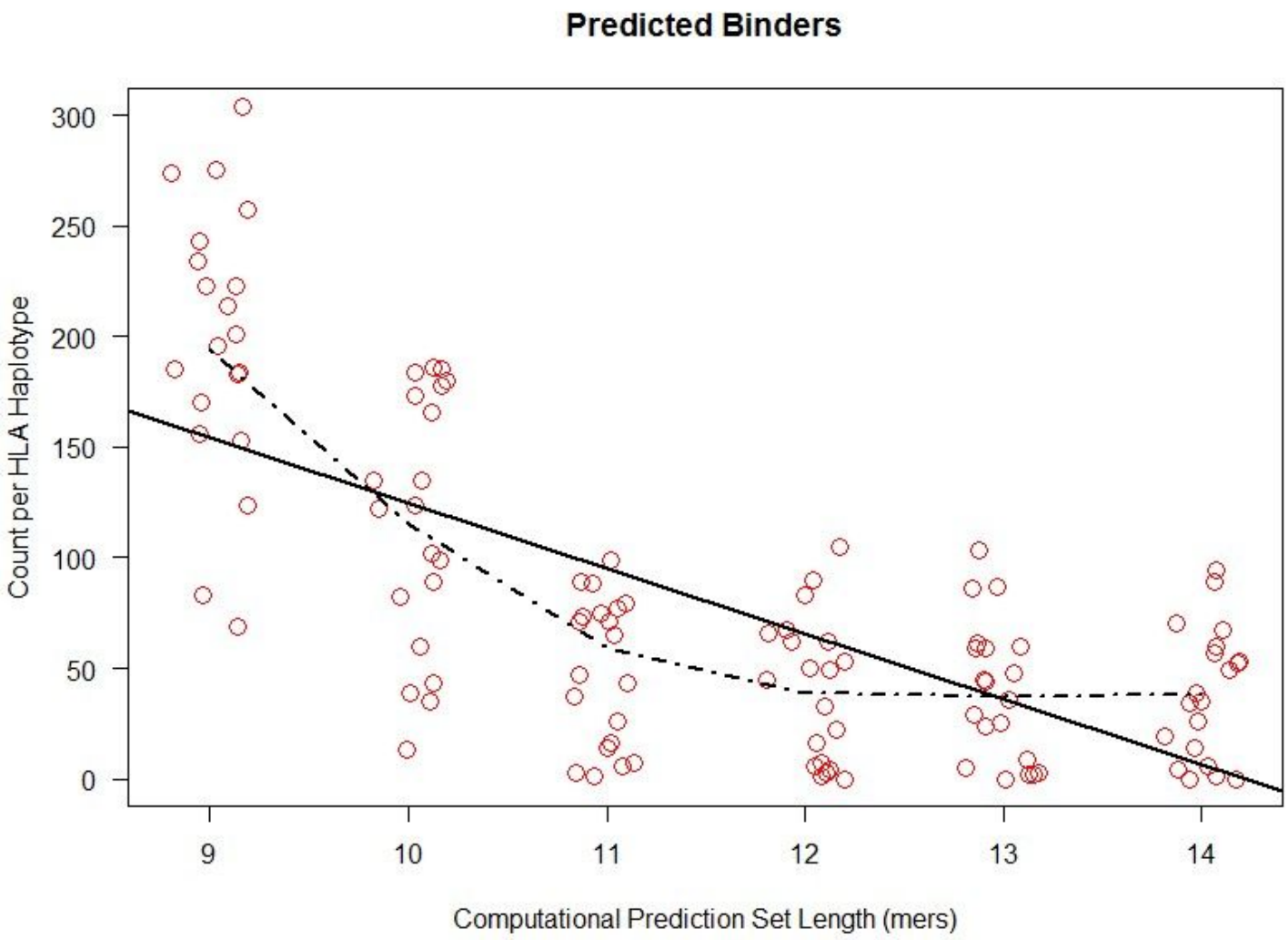

Figure 2

Computational epitope prediction. NetMHCpan4.0 set length plotted against the number of predicted binders per HLA type shows that the number of predictions reduces as the input set length increases. The dotted line is the trend line, whereas the solid line is the line of best fit. The core 9mer epitope sequence anchoring at positions 2 and 9 was similar across the 9mer through 14mer set length except for one 14-mer peptide (hit 72 in table 2).

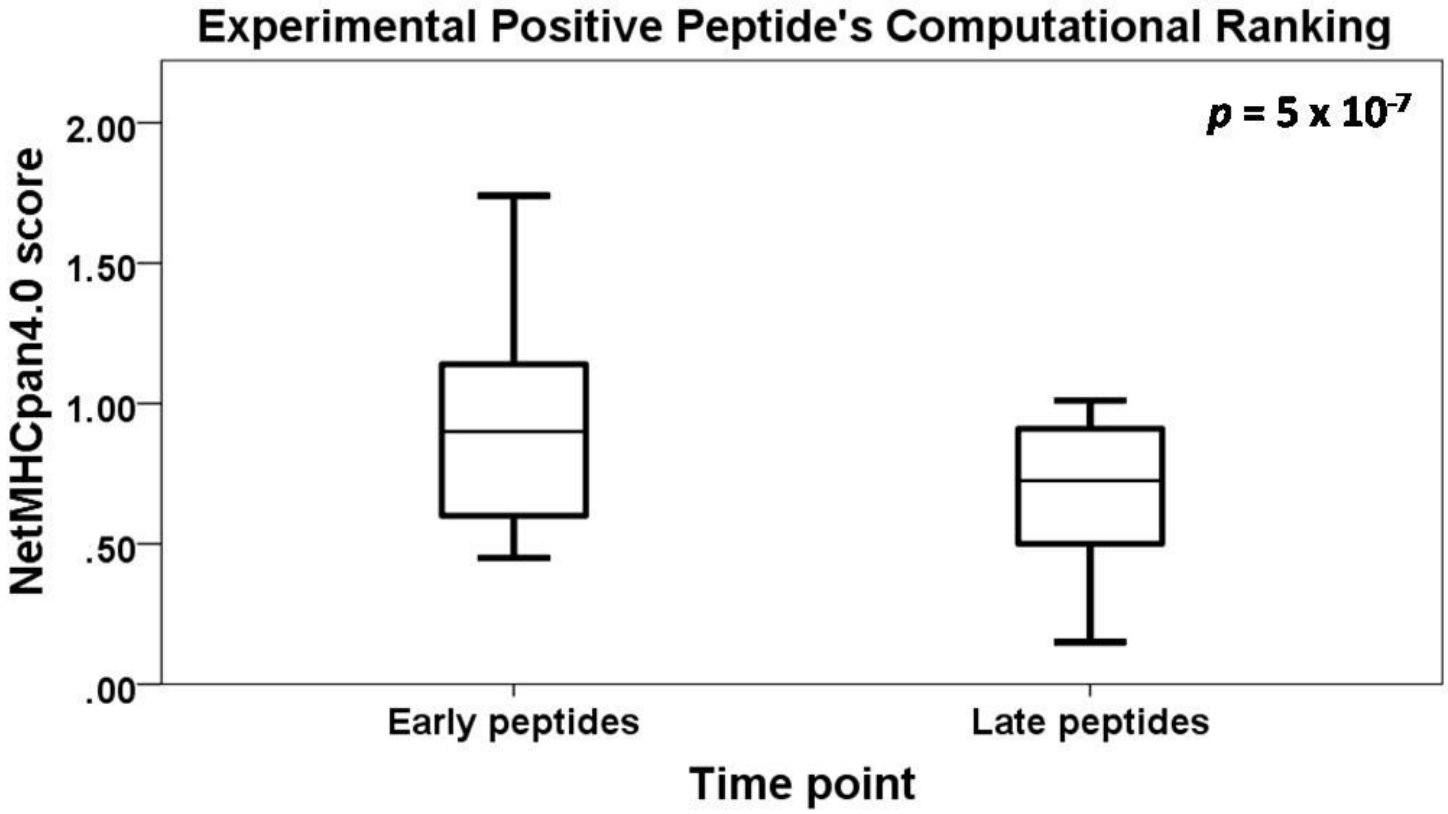

Figure 3

Early versus Late Peptides. Experimentally mapped peptides at baseline (n=34) and at least 12 months later (n=34) were compared using the 9-mer computational NetMHCpan4.0 scores of the hits. The lower the computational score the stronger the predicted binding. Late peptides were significantly stronger binders than early peptides (Wilcoxon signed rank test, p=0.0000005).

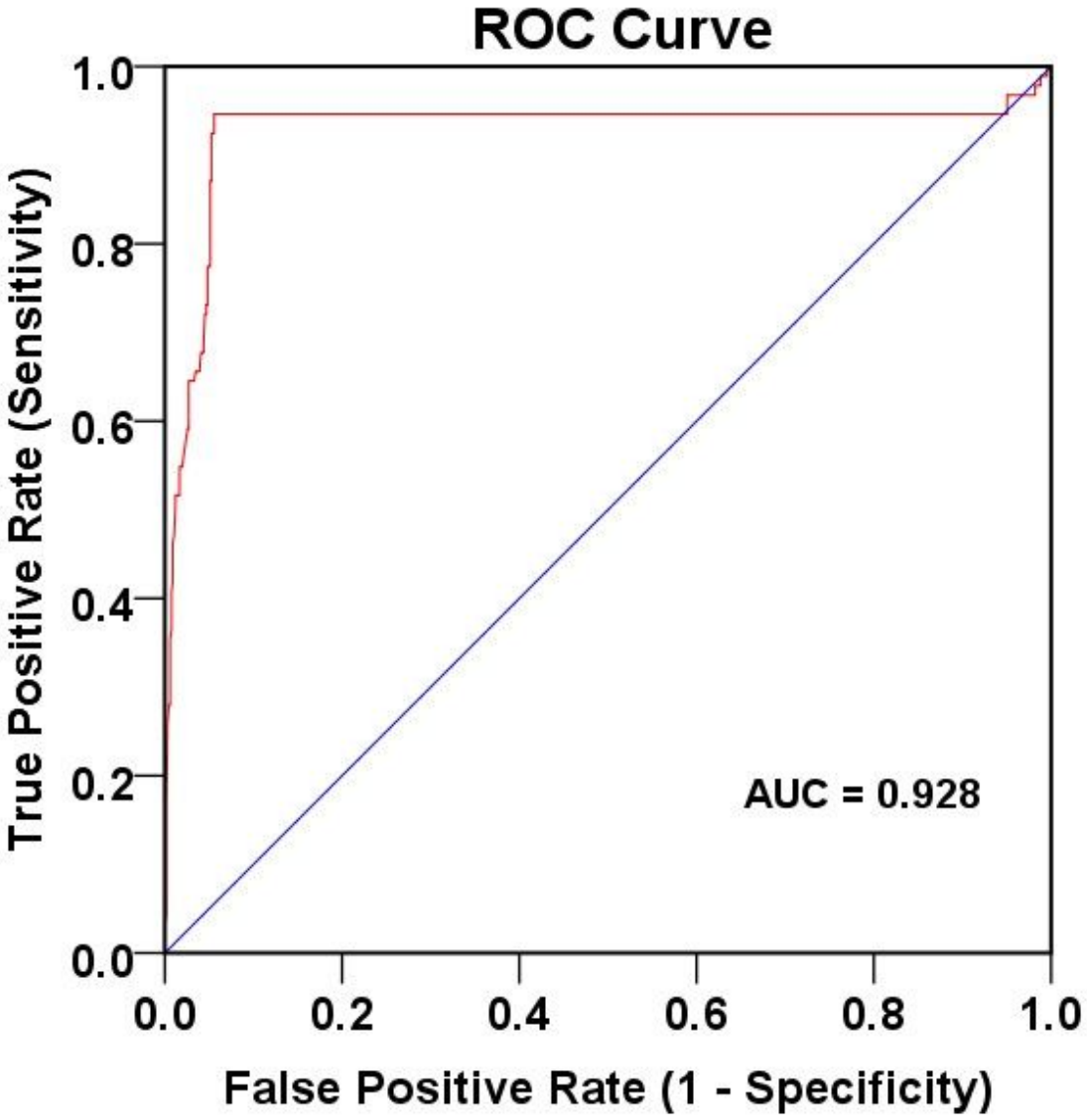

Figure 4

False versus true positive rate for all 9-mer and a single 14-mer test peptides across the 20 test HLA class I types. The diagonal line shows the random guess whereas the red curve shows the observed experimentally mapped epitopes versus the NetMHCpan4.0 expected predictions.

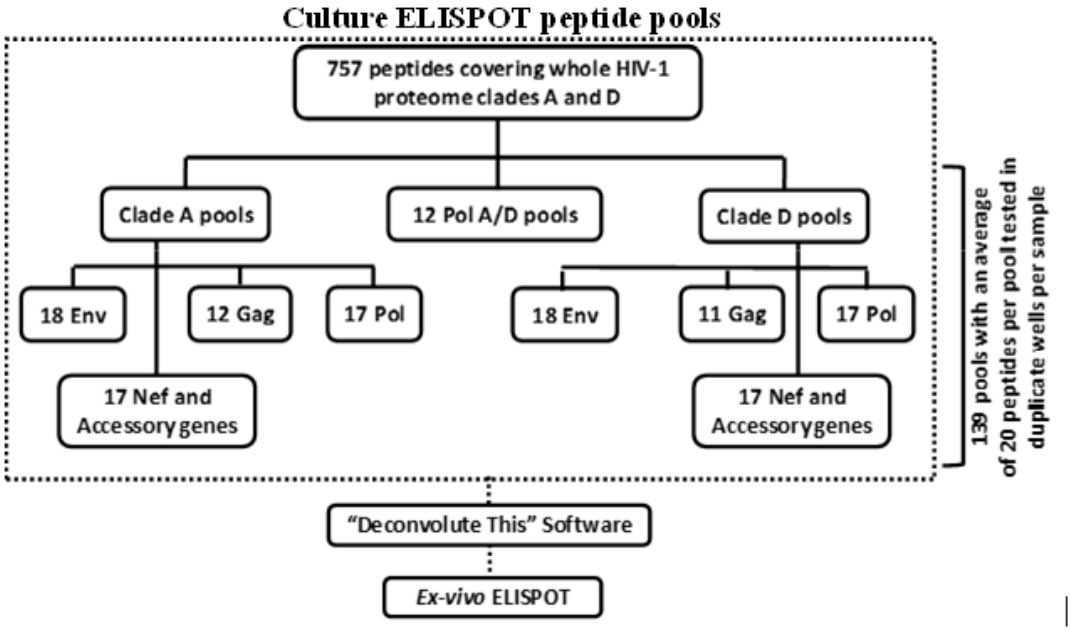

Figure 5

the experimental peptide mapping data was generated culture ELISPOT of multiple peptides pools were tested in duplicate wells per time point followed by ex-vivo ELISPOT of potential candidate epitopes. To experimentally map a single time point required at least 400 assay wells.