

Selective HLA restriction permits the evaluation and interpretation of immunogenic breadth at comparable levels to autologous HLA

Jonathan Hare^{1*}, Andrew Fiore-Gartland², Edward McGowan³, Rachel Rosenthal⁴, Eric Hunter⁵, Jill Gilmour^{1,3} & Morten Nielsen⁶

¹International AIDS Vaccine Initiative, New York, New York, USA

²Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

³IAVI Human Immunology Laboratory, Imperial College, London, UK

⁴Francis Crick Institute, London, UK

⁵Emory Vaccine Center, Atlanta, GA, USA

⁶Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

*Corresponding author jhare@iavi.org

Keywords: HLA diversity, HLA frequency, predicted T-cell epitopes, immunogenic breadth

Abstract

Existing approaches to identifying predictive T-cell epitopes have traditionally utilized either 2-digit HLA super-families or more commonly autologous HLA alleles to facilitate the predictions, but frequently they may not consider their representation within a population. Here we propose a modification to this concept whereby subsets of individuals are selected for their specific HLA allele profiles and the representation they provide within a given population. Using this targeted approach to HLA selection and the linkages to specific individuals may enable the design of restricted experimental strategies.

Introduction

The use of in silico predictive algorithms to identify potential CD8 T-cell epitopes has expanded dramatically in recent years as the number and accuracy of the predictive models increases (Soria-Guerra et al. 2015). A leading algorithm NetMHCpan (Nielsen et al. 2007) utilizes a machine learning approach to HLA/protein interactions to ensure maximum accuracy and has been utilized within a wide variety of disease models for identifying target epitopes (Echigoya et al. 2015; McGranahan et al. 2016).

The traditional approach to utilizing these predictive algorithms has been to include autologous or population MHC Class I haplotypes as a precaution to ensure full HLA coverage. This is predicated on two complimentary assumptions that:

1. The diversity of HLA:epitope presentation within a population cannot be achieved through assessment of a restricted HLA phenotype whilst at the same time maintaining a depth of immunogenic “breadth” for each specific HLA:antigen pair.

2. Immunogenic “breadth” can be empirically assessed as the cumulative number of immunogenic output, where the larger the metric the greater the “breadth”.

In this communication we aim to challenge both assumptions:

Restricted sub-populations represent the HLA diversity of larger cohorts

Assessment of an individuals’ HLA diversity within populations has traditionally been interrogated at two distinct levels; comparison of HLA digits and comparison of HLA haplotype sequence. Both these approaches have some appealing characteristics including intrinsic simplicity and direct linkages to the individuals. However both these approaches also have flaws; at the 4 digit level, two individuals may have similar allele phenotypes (HLA*B57:02 and HLA*B57:03) but can have very different peptide recognition patterns, and therefore represent high epitope presentation diversity (Sidney et al. 2008). At the level of MHC amino acid sequence diversity the reverse applies; two HLA alleles may have very divergent amino acid sequences and yet have very similar binding properties and therefore represent low epitope presentation diversity (Ngumbela et al. 2008).

An alternative metric of population HLA diversity that may better reflect the complexity of the HLA restricted epitope presentation interactions between individuals is to use an epitope predicting algorithm. Utilizing these algorithms, it is possible to generate predicted HLA restricted epitopes and then assess the affinity and stability of each predicted epitope to that HLA (Harndahl et al. 2012). These values can then be represented as a relative measurement of the “distance” between HLA alleles.

69

70 As a theoretical proof of concept for this technique to characterize HLA diversity, a subset of
71 volunteers enrolled within Protocol C, an IAVI HIV epidemiological study (Kamali et al. 2015),
72 were evaluated for their representative frequency and distribution against the full cohort by 4
73 digit characterization and HLA binding profile. The data indicated that at the level of allele
74 frequency, these volunteers were representing >80% of total HLA-A, -B and -C frequency. An
75 HLA binding profile was then computed for each allele by predicting the binding affinity for
76 9mer HIV gag peptides against the binding profile of the total HLA allele distribution from
77 within Protocol C. From these profiles, the overall binding characteristics of each volunteer
78 based on their HLA restriction was compiled. By evaluating the different affinities calculated by
79 the algorithm, it is possible to assign a distance between each allele and each volunteer based on
80 their predicted ability to iteratively bind gag peptides. The distances are then used to cluster
81 alleles/volunteers and visualize HLA diversity (McGowan et al. submitted.). Utilizing these
82 concepts, it is then possible to identify, within a population, individuals with different allele
83 frequencies and distributions that can represent the total HLA diversity within the same
84 population.

85

86 **Re-defining Experimental Immunogenic breadth**

87 A widely-used definition of T-cell restricted immunogenic breadth is that the greater the number
88 of positive responses to discrete linear peptide epitopes, the greater the breadth of the adaptive
89 response. This paradigm functions effectively within the restricted boundaries of an experimental
90 approach, for instance ELISPOT, but struggles once the complexities of multiple individuals,
91 tested against multiple sources of potential experimental target inputs are included, as is the case

in analyzing specific cohorts or populations. To understand immunogenic breadth at the population level requires the linkage of the experimental target inputs of the immunoassay to the population from which they are derived. How this linkage is defined is not always straight forward, although it has been made for some disease models (Seaman et al. 2010) and techniques (Fiore-Gartland et al. 2016). Our interest is in designing and developing an HIV CD8 T-cell restricted candidate vaccine that accounts for the relationships and linkages between an individual's immune repertoire and the prevalent HIV transmission sequences as key parameters to assess to ensure immunogenic breadth.

Traditional assessments of HIV sequence diversity rely on sequence clustering to define breadth, and vaccine strategies utilizing this concept are advancing to clinical evaluation (Ondondo et al. 2016). There are some restrictive assumptions of these approaches however that may undermine the ability of this approach to appropriately reflect breadth, namely:

- They assume that areas of sequence conservation are equally represented in the context of an HLA restricted epitope presentation
- They assume sequence clustering, as a measurement of diversity, accurately represents the immune profile within individuals
- They assume that an empirical determination of immunogenic breadth will capture any linkages of the experimental target input to the population, so that the greater the number of responses the greater the "breadth"
- They equally weight the sequence inputs and make no allowance for the possibility that any sequence may fail to represent the total antigenic breadth

Error! Reference source not found. compares two distance metrics (x and y axis) by plotting the distance by each metric for all pairs of sequences ($N = 130$ sequences). Points that lay on or near the dashed line were similarly distant by the two metrics, while points above or below the line indicate a difference in the metrics.

As expected a distance based on all 9mers is similar to a distance based on 77 random 9mers (**Error! Reference source not found.A**). However, a distance based on the 77 predicted epitopes was systematically shorter than a distance based on all 9mers or on 77 random 9mers (**Error! Reference source not found.B and C**). This suggests that sequences are more similar from the perspective of HLA recognition and potentially from the perspective of the CD8⁺ T cells. By comparing the epitope-based distance to 100 different metrics each based on 77 random epitopes we can compute a p -value for this effect. Out of 100 random sets of epitopes, none produced distances that were shorter than the epitope-based distance on average ($p < 0.01$). This result implies that mutations are generally more common in non-epitopes, i.e. that site-wise entropy should be higher in regions that are not predicted epitopes and therefore should be excluded when determining immunogenic breadth.

Capturing the diversity of the HLA within a population and the subsequent linkages that exist between that population and any associated experimental target inputs are essential for capturing and interpreting immunogenic breadth. By developing quantitative assessments of HLA diversity and relating these same profiles to the immunological relationships of any key disease targets, it would be possible to design restricted experimental strategies that do not introduce bias and can therefore be evaluated as a proxy for the complete population sample.

Conflict of Interest

138 The authors declare that the research was conducted in the absence of any commercial or
139 financial relationships that could be construed as a potential conflict of interest.

140 **Author Contributions**

141 JH wrote the manuscript and provided conceptual input. EM contributed to manuscript and
142 provided conceptual input/ AFG, RR provided technical expertise and contributed to manuscript.
143 EH, JG provided key supervision and support MN developed and implemented the analysis
144 scripts and provided conceptual input.

145 **Acknowledgements**

146 This work was funded in part by IAVI and made possible by the support of the United States
147 Agency for International Development (USAID) and other donors. The full list of IAVI donors is
148 available at <http://www.iavi.org>. The contents of this manuscript are the responsibility of IAVI
149 and do not necessarily reflect the views of USAID or the US Government.

References

- Echigoya, Yusuke et al. 2015. "In Silico Screening Based on Predictive Algorithms as a Design Tool for Exon Skipping Oligonucleotides in Duchenne Muscular Dystrophy" ed. Emanuele Buratti. *PLOS ONE* 10(3): e0120058. <http://dx.plos.org/10.1371/journal.pone.0120058> (February 24, 2017).
- Fiore-Gartland, Andrew et al. 2016. "Pooled-Peptide Epitope Mapping Strategies Are Efficient and Highly Sensitive: An Evaluation of Methods for Identifying Human T Cell Epitope Specificities in Large-Scale HIV Vaccine Efficacy Trials." *PloS one* 11(2): e0147812. <http://www.ncbi.nlm.nih.gov/pubmed/26863315> (August 3, 2016).
- Harndahl, Mikkel et al. 2012. "Peptide-MHC Class I Stability Is a Better Predictor than Peptide Affinity of CTL Immunogenicity." *European Journal of Immunology* 42(6): 1405–16.
- Kamali, Anatoli et al. 2015. "Creating an African HIV Clinical Research and Prevention Trials Network: HIV Prevalence, Incidence and Transmission." *PloS one* 10(1): e0116100. <http://www.ncbi.nlm.nih.gov/pubmed/25602351> (August 3, 2016).
- McGowan, Edward et al. "Utilizing Computational Machine Learning Tools to Understand Immunogenic Breadth in the Context of a CD8 T-Cell Mediated HIV Response." *Submitted*.
- McGranahan, Nicholas et al. 2016. "Clonal Neoantigens Elicit T Cell Immunoreactivity and Sensitivity to Immune Checkpoint Blockade." *Science*.
- Ngumbela, Kholiswa C et al. 2008. "Quantitative Effect of Suboptimal Codon Usage on Translational Efficiency of MRNA Encoding HIV-1 Gag in Intact T Cells." ed. Douglas F. Nixon. *PloS one* 3(6): e2356. <http://dx.plos.org/10.1371/journal.pone.0002356> (January 26, 2017).
- Nielsen, Morten et al. 2007. "NetMHCpan, a Method for Quantitative Predictions of Peptide

173 Binding to Any HLA-A and -B Locus Protein of Known Sequence.” ed. Esper Kallas. *PloS one*
174 2(8): e796. <http://dx.plos.org/10.1371/journal.pone.0000796> (January 29, 2017).

175 Ondondo, Beatrice et al. 2016. “Novel Conserved-Region T-Cell Mosaic Vaccine With High
176 Global HIV-1 Coverage Is Recognized by Protective Responses in Untreated Infection.”
177 *Molecular Therapy* 24(4): 832–42. <http://www.ncbi.nlm.nih.gov/pubmed/26743582> (February
178 27, 2017).

179 Seaman, M S et al. 2010. “Tiered Categorization of a Diverse Panel of HIV-1 Env Pseudoviruses
180 for Assessment of Neutralizing Antibodies.” *J Virol* 84(3): 1439–52.
181 <http://www.ncbi.nlm.nih.gov/pubmed/19939925>.

182 Sidney, John et al. 2008. “HLA Class I Supertypes: A Revised and Updated Classification.”
183 *BMC immunology* 9(1): 1. [http://bmcimmunol.biomedcentral.com/articles/10.1186/1471-2172-9-](http://bmcimmunol.biomedcentral.com/articles/10.1186/1471-2172-9-1)
184 1 (January 26, 2017).

185 Soria-Guerra, Ruth E., Ricardo Nieto-Gomez, Dania O. Govea-Alonso, and Sergio Rosales-
186 Mendoza. 2015. “An Overview of Bioinformatics Tools for Epitope Prediction: Implications on
187 Vaccine Development.” *Journal of Biomedical Informatics* 53: 405–14.

188