Selective HLA restriction enables the evaluation and interpretation of immunogenic breadth at comparable levels to that observed with broader HLA distribution

Jonathan Hare1,2 | Andrew Fiore-Gartland3 | Edward McGowan2 | Eric Hunter4 | Jill Gilmour5 | Morten Nielsen6

1 International AIDS Vaccine Initiative, New York, New York, USA
2 IAVI Human Immunology Laboratory, Imperial College, London, UK
3 Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA
4 Emory Vaccine Center, Emory University, Atlanta, Georgia, USA
5 Department of Infectious Disease, Imperial College, London, UK
6 Department of Health Technology, Technical University of Denmark, Lyngby, Denmark

Correspondence
Jonathan Hare, International AIDS Vaccine Initiative, New York, New York, USA.
Email: jhare@iavi.org

Funding information
United States Agency for International Development

1 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 [1]. A leading algorithm NetMHCpan [2, 3] 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 [4, 5].

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 [6].
2. Immunogenic "breadth" can be empirically assessed as the cumulative number of immunogenic output, where the larger the metric the greater the "breadth" [7].

In this viewpoint we aim to challenge both assumptions:

Abstract

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

KEYWORDS
HLA diversity, HLA frequency, immunogenic breath, predicted T-cell epitopes
1.1 | 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 four 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 [8]. 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 [9].

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 [10]. These values can then be represented as a relative measurement of the “distance” between HLA alleles.

As a theoretical proof of concept for this technique to characterize HLA diversity, a subset of 13 volunteers enrolled within IAVI Protocol C, a longitudinal prospective study of 613 HIV positive volunteers [11, 12], were evaluated for their representative frequency and distribution against the full cohort by 4 digit characterization and HLA binding profile. The data indicated that at the level of allele frequency, these volunteers were representing > 80% of total HLA-A, -B and -C frequency [13]. An HLA binding profile was then computed for each allele by predicting the binding affinity for 9mer HIV gag peptides against the binding profile of the total HLA allele distribution from within Protocol C. From these profiles, the overall binding characteristics of each volunteer based on their HLA restriction were compiled. By evaluating the different affinities calculated by the algorithm, it is possible to assign a distance between each allele and each volunteer based on their predicted ability to iteratively bind gag peptides. The distances are then used to cluster alleles/volunteers and visualize HLA diversity [13]. Utilizing these concepts, it is then possible to identify, within a population, individuals with different allele frequencies and distributions that can represent the total HLA diversity within the same population.

1.2 | Re-defining experimental immunogenic breadth

A widely-used definition of T-cell restricted immunogenic breadth is that the greater the number of positive responses to discrete linear peptide epitopes, the greater the breadth of the adaptive response. This paradigm functions effectively within the restricted boundaries of an experimental approach, for instance ELISPOT, but struggles once the complexities of multiple individuals, 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 straightforward, although it has been made for some disease models [14] and techniques [15]. 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 [16]. 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.

Figure 1 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 (Figure 1A). However, a distance based on the 77 predicted epitopes was systematically shorter than a distance based on all 9mers or on 77 random 9mers (Figure 1B,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, that is, 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.

ACKNOWLEDGMENTS
This work was supported by IAVI and was made possible by generous support from many donors including the Bill and Melinda Gates Foundation, the Ministry of Foreign Affairs of Denmark, Irish Aid, the Ministry of Finance of Japan, the Ministry of Foreign Affairs of the Netherlands, the Norwegian Agency for Development Cooperation (NORAD), the United Kingdom Department for International Development (DFID) and the United States Agency for International Development (USAID). The full list of IAVI donors is available at http://www.iavi.org [17].

CONFLICT OF INTEREST
The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS
J.H. wrote the manuscript and provided conceptual input. E.M. contributed to manuscript and provided conceptual input. AFG provided technical expertise and contributed to manuscript. E.H., J.G. provided key supervision and support. M.N. developed and implemented the analysis scripts and provided conceptual input.
REFERENCES

1. Soria-Guerra, R. E., Nieto-Gomez, R., Govea-Alonso, D. O., & Rosales-Mendoza, S. (2015). An overview of bioinformatics tools for epitope prediction: Implications on vaccine development. Journal of Biomedical Informatics, 53, 405–414.

2. Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., & Buus, S. (2007). NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS One, 2, e796.

3. Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Research, 48, W449–W454.

4. Granahan, N., Furness, A. J. S., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., Jamal-Hanjani, M., Wilson, G. A., Birkbak, N. J., Hiley, C. T., Watkins, T. B. K., Shafi, S., Murugaeus, N., Mitter, R., Akara, A. U., Linares, J., Marafioti, T., Henry, J. Y., Van Allen, E. M., Miao, D., Schilling, B., Schadendorf, D., Garraway, L. A., Makarov, V., Rizvi, N. A., Snyder, A., Hellmann, M. D., Mergelbaur, T., Wolchok, J. D., Shukla, S. A., Wu, C. J., Peggs, K. S., Chan, T. A., Hadrup, S. R., Quezada, S. A. & Sivamurthy, S. (2016). Clonal neoantigens elicit T-cell immunity in individuals with broader HLA distribution. Science, 351, 1463–9.

5. Echigoya, Y., Mouly, V., Garcia, L., Yokota, T., & Duddy, W. (2015). in silico screening based on predictive algorithms as a design tool for exon skipping oligonucleotides in Duchenne muscular dystrophy. PLoS One, 10, e0120058.

6. Sette, A. & Sidney, J. (1999). Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism. Immuno- genetics, 50, 201–212.

7. Manuel, E. R., Charini, W. A., Sen, P., Peyerl, F. W., Koroda, M. J., Schmitz, J. E., Autissier, P., Sheeter, D. A., Torbett, B. E., & Letvin, N. L. (2006). Contribution of T-cell receptor repertoire breadth to the dominance of epitope-specific CD8+ T-lymphocyte responses. Journal of Virology, 80, 120–32 4040.

8. Sidney, J., Peters, B., Frahm, N., Brander, C., & Sette, A. (2008). HLA class I supertypes: A revised and updated classification. BMC Immunology, 9, 6.

9. Ngumbela, K. C., Ryan, K. P., Sivamurthy, R., Brockman, M. A., Gandhi, R. T., Bhardwaj, N., & Kavanaugh, D. G. (2008). Quantitative effect of suboptimal codon usage on translational efficiency of mRNA encoding HIV-1 gag in intact T cells. PLoS One, 3, e2356.

10. Harndahl, M., Rasmussen, M., Roder, G., Dalgaard Pedersen, I., Sørensen, M., Nielsen, M., & Buus, S. (2012). Peptide-MHC class I stability is a better predictor than peptide affinity of CTL immunogenicity. European Journal of Immunology, 42, 10405–10416.

11. Kamali, A., Price, M. A., Lakhi, S., Karita, E., Inambao, M., Sanders, E. J., Anzala, O., Latka, M. H., Bekker, L.-G., Kaleebu, P., Asiki, G., Ssetaala, A., Ruzagira, E., Allen, S., Farmer, P., Hunter, E., Mutua, G., Makkan, H., Tichacek, A., Brill, I. K., Fast, P., Stevens, G., Chetty, P., Amornkul, P. N., & Gilmour, J. (2015). Creating an African HIV clinical research and prevention trials network: HIV prevalence, incidence and transmission. PLoS One, 10, e0116100.

12. Amornkul, P. N., Karita, E., Kamali, A., Rida, W. N., Sanders, E. J., Lakhi, S., Price, M. A., Kilembe, W., Cormier, E., Anzala, O., Latka, M. H., Bekker, L.-G., Allen, S., Gilmour, J., & Fast, P. E. (2013). Disease progression by infecting HIV-1 subtype in a seroconverter cohort in sub-Saharan Africa. AIDS, 27, 2775–2786.

13. Mcgowan, E., Rosenthal, R., Fiore-Gartland, A., Macharia, G., Balinda, S., Kapaata, A., Umvilighozo, G., Muok, E., Dalel, J., Streetfield, C. L., Coutinho, H., Dillernia, D., Monaco, D. C., Morrison, D., Yue, L., Hunter, E., Nielsen, M., Gilmour, J., & Hare, J. (2021). Utilizing computational machine learning tools to understand immunogenic breadth in the context of a CD8 T-cell mediated HIV response. Frontiers in Immunology, 12, 609884.

14. Seaman, M. S., Janes, H., Hawkins, N., Grandpre, L. E., Devoy, C., Giri, A., Coffey, R. T., Harris, L., Wood, B., Daniels, M. G., Bhattacharya, T., Lapedes, A., Polonis, V. R., Mccutchan, F. E., Gilbert, P. B., Self, S. G., Korber, B. T., Montefiori, D. C., & Mascola, J. R. (2010). Tiered categorization of a diverse panel of HIV-1 Env pseudoviruses for assessment of neutralizing antibodies. Journal of Virology, 84, 1439–1452.

15. Fiore-Gartland, A., Manso, B. A., Friedrich, D. P., Gabriel, E. E., Finak, G., Moodie, Z., Hertz, T., De Rosa, S. C., Frahm, N., Gilbert, P. B., & Mccllrath, M. J. (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, e0147812.

16. Ondondo, B., Murakoshi, H., Clutton, G., Abdul-Jawad, S., Wee, E. G.-T., Gatanaga, H., Oka, S., Mcmichael, A. J., Takiguchi, M., Korber, B. T., Montefiori, D. C., & Mascola, J. R. (2013). Disease progression of an HLA-B27+ subject in the PEPFAR trial and highly sensitive: An evaluation of methods for identifying human T cell epitope specificities in large-scale HIV vaccine efficacy trials. PLoS One, 9, 609884.

17. Hare, J., Fiore-Gartland, A., Mcgowan, E., Rosenthal, R., et al. (2020). Selective HLA restriction permits the evaluation and interpretation of immunogenic breadth at comparable levels to autologous HLA.

How to cite this article: Hare, J., Fiore-Gartland, A., Gowan, E. M., Hunter, E., Gilmour, J., Nielsen, M. (2021). Selective HLA restriction enables the evaluation and interpretation of immunogenic breadth at comparable levels to that observed with broader HLA distribution. Proteomics, 21, e2100121. https://doi.org/10.1002/pmc.202100121