Quantification of heterogeneity in CD8$^+$ T cell responses to vaccine proteins: an MHC-guided perspective

Duane Harris, Ruy M. Ribeiro, Alan S. Perelson, Ben Mcmahon, and Carmen Molina-París

Abstract Vaccines exploit the exceptional ability of the adaptive immune system (IS) to respond to, and remember, encounters with viruses. Novel vaccine technologies (*e.g.,* viral vector, DNA, or RNA) enable a "plug and play" approach to *immunogen* (part of the virus that can be recognized by the IS) design. These technical advances inherently raise a number of questions in vaccine immunology. First, the genetic diversity of highly variable viruses makes it difficult to identify an immunogen that can be used to vaccinate against infection. Second, the most effective route to vaccine efficacy and protection is to engage multiple arms of the IS. Most current vaccine strategies aim to confer protection with antibodies (humoral response), which are produced by B cells. There is also substantial evidence of *protective cellular immunity* correlated with CD8$^+$ T cell responses to *conserved regions* of the genome of HIV-1, Lassa, and Ebola virus. Yet we do not currently have the ability to engineer vaccines that exploit T cell responses. Enabling this second arm to engage the adaptive immune system to fight infections would be a game changer. In this talk, I shall discuss quantitative methods we have developed to evaluate, quantify, and predict T cell responses to a broad class of viral infections that will support the rational design of T cell-based vaccines. The following biological insight supports our effort: cytotoxic (CD8$^+$) T cells kill those cells they recognize as being infected. To identify infected cells, CD8$^+$ T cells work in tandem with the infected cell: the infected cell can "cut" viral proteins and expose short protein fragments (peptides about 9 amino acid long) on designated signposts on its surface, and the T cell looks for these signposts; if it "sees" an "offending" viral peptide, it starts the process of killing the infected cell. Human genetics is involved in the types and quality of available signposts (the technical term is the major histocompatibility complex (MHC)) displayed on the surface of infected cells. This, in great part, explains the large variability in human immune responses to vi

Figure 1: Each of us has six different MHC class I molecules: two of each type, HLA-A, HLA-B, and HLA-C.

Our approach recognizes three important sources of variability. First, human MHC molecules are divided into three major types (HLA-A, HLA-B, and HLA-C, see Figure 1), which can be further subdivided into MHC subtypes. Second, T cell binding strength (recognition) of peptides depends both on the (viral) peptide and the MHC subtype. Finally, and third, geographically distinct human populations have specific and characteristic distributions of prevalent MHC subtypes. Our innovation in building a predictive model for T cell-based vaccines is to encode these three factors making use of a tri-partite graph (see Figure 2). Our work leverages existing genome-wide association studies between host and virus genomes showing that MHC selective pressure drives *in vivo* viral evolution (*e.g.,* HCV and HIV-1), as well as protection (*e.g.,* HIV-1 control and in malaria-endemic environments). The tri-partite graph, $\mathcal{G}$, links eleven geographical regions $\mathcal{R} = \{r_1, r_2, \ldots, r_K\}$ covering the world's human population, M MHC alleles $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}$, and N conserved 9-mer peptides from selected viral proteins $\mathcal{P} = \{p_1, p_2, \ldots, p_N\}$. Edges between nodes from different sets connect *i)* a geographical region and an MHC allele to encode the frequency of that allele in the region, *i.e.,* $f_3^{(1)}$ is the frequency in r_1 of allele a_3; and *ii)* an MHC allele and a viral peptide to encode the binding affinity between the MHC and the peptide, *i.e.,* s_{51} is the binding score of allele a_5 to peptide p_1 (see Figure 2). This novel graph approach allows us to address the above problems: *1)* the viral genetic diversity is encoded in the set of peptides, $\mathcal{P}$, so that wild type virus and all circulating (or predicted) variants can be analyzed, and *2)* MHC heterogeneity is considered via geographical regions $\mathcal{R}$, MHC alleles $\mathcal{A}$, and their frequencies. $\mathcal{G}$ straightforwardly provides a *metric* to quantify the *regional vaccine coverage* provided by T cell responses, and the framework to unravel the role of worldwide MHC diversity to maximize the *individual vaccine coverage*.

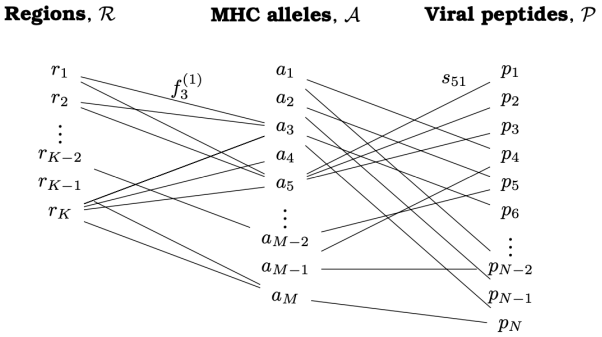

Figure 2: CD8$^+$ T cell immunogen recognition as a tri-partite graph, $\mathcal{G}$. Only a subset of the edges is shown for clarity.

Reference Our methods and approaches have been published in *Quantification of heterogeneity in human CD8$^+$ T cell responses to vaccine antigens: an HLA-guided perspective*, Harris *et al.*, Frontiers in Immunology, **15**, 2024. This manuscript has been reviewed at Los Alamos National Laboratory and assigned report numbers LA-UR-24-23493 and LA-UR-24-30663.

Short bio Carmen Molina París is a Theoretical Physicist (undergraduate at the Universidad de Granada and PhD at the University of Texas at Austin). Her research has transitioned from Theoretical Physics to Theoretical Immunology and Virology. In doing so she created and established a research group in the School of Mathematics at the University of Leeds, that successfully led and coordinated four international EU networks (FP7 and H2020), as well as collaborations with some of the best immunologists and virologists in the world. She has contributed to defining and developing novel mathematical models of relevance to T cell immunology and to human pathogenic infection. She has led research training and educational initiatives in Quantitative T cell Immunology (QuanTI and QuanTII, two EU funded research and training networks). She is a senior scientist in the Theoretical Division at Los Alamos National Laboratory.