



Influence of HLA Class II Polymorphism on Predicted Cellular Immunity Against SARS-CoV-2 at the Population and Individual Level

Hannah C. Copley^{1,2}, Loren Gragert^{3,4}, Andrew R. Leach² and Vasilis Kosmoliaptis^{1,5,6*}

¹ Department of Surgery, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom, ² European Bioinformatics Institute (EMBL-EBI), Cambridge, United Kingdom, ³ Department of Pathology, Tulane University School of Medicine, New Orleans, LA, United States, ⁴ Bioinformatics Research, National Marrow Donor Program, Minneapolis, MN, United States, ⁵ National Institute of Health Research (NIHR) Blood and Transplant Research Unit in Organ Donation and Transplantation, University of Cambridge, Cambridge, United Kingdom, ⁶ NIHR Cambridge Biomedical Research Centre, Cambridge, United Kingdom

OPEN ACCESS

Edited by:

Yongqun Oliver He,
University of Michigan,
United States

Reviewed by:

Arman Bashirova,
Leidos Biomedical Research, Inc.,
United States

Jennifer Ann Juno,
The University of Melbourne,
Australia

Grigory Efimov,
National Research Center for
Hematology, Russia

*Correspondence:

Vasilis Kosmoliaptis
vk256@cam.ac.uk

Specialty section:

This article was submitted to
Viral Immunology,
a section of the journal
Frontiers in Immunology

Received: 18 February 2021

Accepted: 28 June 2021

Published: 19 July 2021

Citation:

Copley HC, Gragert L, Leach AR
and Kosmoliaptis V (2021) Influence
of HLA Class II Polymorphism
on Predicted Cellular Immunity
Against SARS-CoV-2 at the
Population and Individual Level.
Front. Immunol. 12:669357.
doi: 10.3389/fimmu.2021.669357

Development of adaptive immunity after COVID-19 and after vaccination against SARS-CoV-2 is predicated on recognition of viral peptides, presented on HLA class II molecules, by CD4+ T-cells. We capitalised on extensive high-resolution HLA data on twenty five human race/ethnic populations to investigate the role of HLA polymorphism on SARS-CoV-2 immunogenicity at the population and individual level. Within populations, we identify wide inter-individual variability in predicted peptide presentation from structural, non-structural and accessory SARS-CoV-2 proteins, according to individual HLA genotype. However, we find similar potential for anti-SARS-CoV-2 cellular immunity at the population level suggesting that HLA polymorphism is unlikely to account for observed disparities in clinical outcomes after COVID-19 among different race/ethnic groups. Our findings provide important insight on the potential role of HLA polymorphism on development of protective immunity after SARS-CoV-2 infection and after vaccination and a firm basis for further experimental studies in this field.

Keywords: human leukocyte antigens, SARS-CoV-2, T-cells, cellular immunity, COVID-19

INTRODUCTION

The new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for coronavirus disease 2019 (COVID-19), has caused an ongoing pandemic with 93,805,612 confirmed cases and 2,026,093 deaths worldwide [as of 19 January 2021 (1)]. Several risk factors for severe COVID-19 are now well-established, including age, gender, obesity and various comorbidities such as diabetes, cancer and cardiovascular or chronic lung disease (2–5). There is, however, an urgent need to better understand the role of race and ethnic differences on health outcomes. Several studies have highlighted a disproportionate prevalence of COVID-19 infections, higher rates of hospitalisation, and increased incidence of death in people from black and minority ethnic groups but the underlying reasons for these observations are not well understood (2, 6–8).

Even after accounting for well-known risk factors, there is still wide inter-individual clinical variability of COVID-19 outcomes within considered risk groups, which may reflect underlying genetic differences (9). The principal genetic region involved in immunity against viral pathogens is the Major Histocompatibility Complex encompassing the Human Leukocyte Antigen (HLA) loci. HLA class I (HLA-A, -B, -C) and HLA class II (HLA-DR, -DQ, -DP) proteins present viral peptides for recognition by CD8⁺ and CD4⁺ T-cells, respectively. The latter orchestrate adaptive anti-viral immunity and drive B-cell activation and maturation for robust humoral responses. Extensive polymorphism is observed in the HLA system, resulting in differences in HLA allele frequency both within and across human populations. HLA genotype can be a determining factor in development of protective immunity and, in turn, may account for part of the observed heterogeneity in measured immune responses and in clinical outcomes after SARS-CoV-2 infection (10, 11). It is also well established that there is marked biological variation in how individuals respond and maintain immunity after vaccination which is, in part, attributable to genetic factors (12). It is, therefore, important to consider the role of HLA polymorphism when designing viral subunit or peptide vaccine formulations and in assessing population coverage and likelihood of immune protection after vaccination (13).

When considering the role of the HLA system in COVID-19 susceptibility and vaccine responses, it is essential to account for differences in HLA allele frequencies across human populations and, importantly, for the linkage disequilibrium between HLA loci that result in population-specific haplotype frequencies. Here, we utilise information on human HLA haplotype frequencies of twenty five human populations (four broad population categories and twenty one detailed population subcategories) at an unprecedented scale, capitalising on the extensive high-resolution HLA data deposited in the National Marrow Donor Program Registry, to compute population level immune responses against SARS-CoV-2 based on predicted high-affinity binding of viral proteome derived peptides by HLA class II molecules. Overall, we find similar potential for anti-SARS-CoV-2 cellular immunity across all populations examined suggesting that HLA polymorphism is unlikely to account for observed disparities in clinical outcomes after COVID-19 among different race and ethnic groups. However, within populations, we identify wide variability among individuals in predicted CD4⁺ T-cell reactivity against structural, non-structural, and accessory SARS-CoV-2 proteins, according to HLA genotype. Nevertheless, we predict robust immune reactivity against the SARS-CoV-2 Spike protein, the basis for the majority of current vaccination efforts, both at the population and at the individual level.

MATERIALS AND METHODS

Identification of Potential T-Cell Epitopes

Full viral proteome sequences for SARS-CoV-2 were downloaded from UniProt (14). FASTA-formatted protein

sequence data for each protein and protein class were examined individually and in combination. We produced potential peptides of 15 amino acids length (15mers), using sliding windows over the entire proteome. Proteins of fewer than 15 amino acids in length were not examined. Analyses were performed considering proteins individually, some protein domains individually, and in groupings of proteins (both the whole proteome, and all structural, non-structural, and accessory proteins).

HLA and Haplotype Frequency Computation

HLA population frequencies were obtained from US unrelated stem cell donor registry National Marrow Donor Program (NMDP)/Be The Match. High resolution HLA Class II haplotype frequencies (DRB1, DRB3/4/5, DQA1, DQB1, DPA1 and DPB1 loci) were estimated using an expectation-maximization algorithm [as described by Gragert et al., 2013 (15)] utilising a cohort of 8.9 million US volunteer donors (NMDP/Be The Match registry snapshot 29/05/2020) HLA typed by molecular methods (**Table S1** showing number of individuals genotyped). US population categories were developed based on a race/ethnicity questionnaire included on the donor consent form. There were four broad population categories (European American, African American, Asian or Pacific Islander and Hispanic) and 21 detailed populations subcategories (African American, African Black, South Asian Indian, American Indian - South or Central American, Alaska native of Aleut, North American Indian, Caribbean Black, Caribbean Hispanic, Caribbean Indian, European Caucasian, Filipino, Hawaiian or other Pacific Islander, Japanese, Korean, Middle Eastern or North Coast of Africa, Mexican or Chicano, Chinese, Hispanic - South or Central American, Black - South or Central American, Southeast Asian, Vietnamese). Some individuals were assigned only membership of a broad population category.

Following Hardy-Weinberg equilibrium proportions, multi-locus HLA Class II genotypes were generated by randomly sampling two haplotypes from the same population HLA haplotype frequency distribution. Simulated genotypes were generated for each of the four broad and 21 detailed population groups, with ten replicates at each population of size 1,000, 5,000 and 10,000.

For population-based analyses at the haplotype level, we analysed haplotypes up to 99% cumulative coverage within each population. HLA alleles, when examined individually, included all HLA alleles present in any of these selected haplotypes for the four broad and 21 detailed population groups.

Predicted T-Cell Epitope Identification

Peptide binding affinity was assessed for all Class II HLA that featured in haplotypes in any of the ethnic populations studied using NetMHCIIpan v4.0 (16). Peptides were examined both for their predicted binding affinity (nM) and their percentage rank (compared to a pool of representative peptides for the corresponding HLA). Peptides were defined as binders if their

binding affinity was equal to or less than 500nM and their percentage rank was equal to or greater than 2% (default parameter for strong HLA class II peptide binders). Within the 9,590 15-mer peptides in the entire SARS-CoV-2 proteome, 4,289 peptides were predicted to bind strongly to one or more HLA Class II molecule. Analyses were also performed using alternative binding threshold cut-offs of ≤ 50 nM peptide binding affinity threshold and $\leq 0.5\%$ percentage rank in combination with a ≤ 500 nM peptide binding affinity threshold. The NetMHCIIpan-4.0 programme also identifies the predicted 9-mer binding core. Of all HLA observed, on average 0.52% of haplotypes contained an HLA that was not in the list of 5,620 HLA available to run on NetMHCIIpan-4.0. Total predicted peptide counts for an individual HLA per protein or per whole proteome were calculated by counting peptides only if they were both unique to one another (i.e., a unique 15-mer), but also if the predicted binding core (9-mer) was also unique. This prevented the count from appearing falsely elevated due to sequential overlapping peptides which were presenting the same core, by preventing them from being counted more than once. Total peptide counts (in the context of an HLA haplotype or genotype) were also enumerated from peptides with both unique 15-mers and 9-mer cores, from binding peptides presented in any Class II HLA within the haplotype or genotype.

Quantification and Statistical Analysis

Peptide-HLA scoring models were assessed using scikit-learn (17) in Python, specifically using `sklearn.metrics.roc_auc_score` (AUROC), `sklearn.metrics.average_precision_score` (Average Precision), `sklearn.metrics.accuracy_score` (Accuracy), and `sklearn.metrics.classification_report` (Sensitivity and Specificity) functions (see **Table S3**, with $n=93$ for peptide-HLA combinations). In order to assess AUROC for two predictor variables, a composite predictor was calculated consisting of the normalised sum of both. Population comparisons of peptide scores were performed by calculating the mean and standard deviation using NumPy in Python (see **Table S4**, $N=10,000$ individuals in simulated population groups). Population simulations were assessed for normality using `statistics.shapiro` (Shapiro-Wilk test) and following demonstration of non-normality for all distributions examined, with `statistics.kruskal` (Kruskal-Wallis one-way analysis of variance) to assess differences between replicates of population simulations (all $N=10,000$), both implemented in Python.

RESULTS

Experimental Approach

The principal objective of our study was to examine whether genetic variation at the HLA loci may influence immune responses, and therefore COVID-19 clinical outcomes or response to vaccination, among patients from different ethnic populations. HLA class II molecules present peptides from exogenous antigens for $CD4^+$ T-cell recognition and are therefore critical components for an effective adaptive immune

response that incorporates humoral (B-cell) and cytotoxic ($CD8^+$ T-cell) arms.

We analysed peptide binding for all classical HLA class II loci (HLA-DRB1, -DRB3/4/5, -DQA1, -DQB1, and -DPA1, -DPB1). The main analysis focused on four broad population categories, African Americans (AFA), European Caucasians (CAU), Hispanics (HIS) and Asian/Pacific Islanders (API) and sub-analysis included 21 detailed population subgroups of these broad categories. To account for linkage disequilibrium between HLA class II loci, we analysed HLA haplotypes rather than assuming allele frequencies were independent at each locus. HLA haplotype frequencies were estimated utilising HLA genotype data obtained from a cohort of approximately 8.9 million donors (NMDP/Be The Match registry; **Table S1**). We analysed a total of 25,128 unique haplotypes to enable 99% coverage of each population which in turn required the analysis of 803 HLA class II molecules (see **Supplemental Information**). **Figure S1** depicts the number of distinct HLA-DRB1, -DRB3/4/5 alleles and HLA-DQA1/DQB1, DPA1/DPB1 heterodimers examined and associated haplotype coverage in each population. To confirm the robustness of our observations at the individual level, we analysed multiple replicate genotype samples and demonstrate with replicate sets of simulated genotypes of 10,000 individuals that HLA diversity, as measured by the alpha parameter fits to the power law distribution, was stable, and the number of unique haplotypes needed to reach 95% cumulative frequency was also stable across replicates (**Table S6**).

Immunogenic T-cell epitopes were identified using the full-length reference SARS-CoV-2 sequence (18) with 9,744 amino acids, which was subdivided into the four structural proteins (E, M, N, and S) and 7 additional open reading frames encoding non-structural proteins (NSP1-16) and accessory proteins (proteins 3a, 6, 7a, 7b, 8 and 10). A sliding window of 15 amino acids length was used and, to minimise redundancy, peptides were only counted towards totals if the HLA class II binding core was unique. To evaluate peptide binding and stable display by HLA class II molecules we employed NetMHCIIpan-4.0 which outputs predicted peptide-HLA binding affinity (IC_{50}) in nanomolar units and percentile rank of binding affinity compared to a set of 100,000 random natural peptides. The percentile rank enables incorporation of information from the biological antigen presentation pathway in addition to the peptide-MHC binding event (16, 19). We aimed to detect strongly binding peptides and, therefore, elected to use a threshold of $\leq 2\%$ percentile rank (default parameter for HLA class II strong peptide binding) combined with an affinity threshold ≤ 500 nM (16). To further increase the stringency of criteria for prediction of peptide binding, such that we maximise precision and reduce the rate of false positive peptides (accepting we may not recall all possible peptides that could be displayed), we also explored a model involving a threshold of $\leq 0.5\%$ percentage rank and ≤ 500 nM binding affinity. Finally, we examined a binding affinity threshold of ≤ 50 nM, as previously suggested (13). To validate the computational models we analysed publicly available datasets of experimentally

determined, immunogenic SARS-CoV-2 peptides. We focused on the two largest datasets recently described by Snyder et al. (20, 21) and Nolan et al. (22) (250 HLA class II peptides) and by Mateus et al. (23) (135 HLA class II peptides) that contain peptides from the entire SARS-CoV-2 proteome. We also examined two relatively small datasets encompassing 9 nucleocapsid and 25 structural protein-derived (Spike, nucleocapsid or membrane) peptides (24, 25). In the aforementioned datasets, the HLA restriction was not known and validation focused on positive identification of experimental peptides by the computational models (**Supplemental Information**). We also used an HLA-DRB1*04:01 restricted peptide dataset determined using an *in vitro* peptide-HLA stability assay (26). Our analyses showed that scoring peptide binding based on a combination of $\leq 2\%$ percentile rank and $\leq 500\text{nM}$ binding affinity achieved the best true positive rate (sensitivity) for predicting experimentally derived SARS-CoV-2 peptides (**Table S2**). Similarly, in the HLA restricted dataset by Prachar et al, the combined $\leq 2\%$ percentile rank and $\leq 500\text{nM}$ binding affinity threshold had an AUROC of 0.85 and provided the best combination of precision and specificity in classifying stable peptide binders compared to alternative scoring methods (**Table S3A**). Finally, we validated our approach using a recently published dataset of experimentally determined CD4⁺ T-cell epitopes (27) and demonstrated that our approach does not introduce systematic bias across the studied populations and does not over-predict the number of immunogenic epitopes (**Table S3B**). It should be noted that using a 50nM threshold, without accounting for binding characteristics of natural peptides, resulted in 46% of HLA alleles examined lacking presentation of any SARS-CoV-2 peptides and a bias towards HLA-DR as the major SARS-CoV-2 peptide presenting molecules (**Figure S2**). Nevertheless, we have confirmed key aspects of our analyses using both a $\leq 0.5\%$ percentage rank in combination with a $\leq 500\text{nM}$ peptide binding affinity threshold, and using a $\leq 50\text{nM}$ peptide binding affinity threshold alone. Overall, a total of 9,590 15-mer peptides, derived from all 11 SARS-CoV-2 genes, were examined and 4,289 peptides were predicted to bind strongly according to our defined criteria to at least one HLA class II molecule.

SARS-CoV-2 Viral Proteome Presentation at the Molecular HLA Class II Level

We first examined presentation of viral epitopes by all HLA class II molecules contained in haplotypes representing 99% of the four major broad ethnic populations. Assessment of presentation capacity at the entire viral proteome level (**Figure 1A**) showed that the majority of HLA class II molecules are capable of presenting SARS-CoV-2 peptides, albeit with significant variability. HLA-DR alleles have the highest viral peptide presentation capacity followed by HLA-DP and, to a significantly lower extent, HLA-DQ molecules. Notably, certain common individual HLA molecules were predicted to have very limited ability to present viral peptides, including DQA1*03:01~DQB1*02:01 (no peptide presentation from any protein within the SARS-CoV-2 proteome; with frequency of

3.2% within AFA haplotypes, <0.01% within API haplotypes, 0.1% within CAU haplotypes and 0.5% within HIS haplotypes), DRB1*03:02 (two peptides presented in total from the entire proteome; with frequency of 6.3% within AFA haplotypes, 0.01% within API haplotypes, 0.04% within CAU haplotypes and 1.0% within HIS haplotypes). Consistent with its known high immunogenicity (10, 27–31), Spike protein derived peptides showed strong binding for the majority of HLA class II molecules although presentation capacity again varied and was lowest (no peptides presented) for relatively common alleles such as DQA1*01:01~DQB1*05:03, with a frequency of 1.9–5.4% in each of the four broad population groups (**Figure 1B**). This observation was more prominent for Nucleocapsid derived peptides where strong binding was predicted to be absent in 117 out of 306 HLA molecules found in 99% of haplotypes in the four broad ethnic populations (mostly reflecting HLA-DP molecules and to some extent -DQ molecules; **Figure 1C**). Similar findings were noted for the relatively small Membrane and Envelope structural proteins, with the latter predicted to be non-immunogenic for the majority of common HLA class II molecules (**Figures 1D, E**). For non-structural proteins, HLA class II presentation capacity was variable and dependent upon protein amino acid sequence length (data not shown).

SARS-CoV-2 Proteome Immunogenicity at the Population Level

The capacity of individuals to present viral peptides for recognition by CD4⁺ T-cells depends on the composition of their HLA class II alleles in their inherited haplotypes. Given variation in population-specific HLA haplotype frequencies, we hypothesised that potential differences in SARS-CoV-2 proteome immunogenicity (as reflected by HLA class II peptide presentation) at the population level may reflect disparities in capacity for effective anti-SARS-CoV-2 immunity which could in turn influence response to vaccination or underpin observed variability in COVID-19 clinical outcomes among different ethnic populations. To investigate this hypothesis, we examined viral proteome presentation by HLA class II, accounting for the distribution of HLA haplotypes in a population. This analysis showed that the overall capacity for HLA peptide presentation, at the whole SARS-CoV-2 proteome level, among the four broad ethnic populations examined was remarkably similar (**Figure 2A**). This was also the case considering T-cell epitopes from each SARS-CoV-2 protein (**Table S4**), suggesting that polymorphism at the HLA genomic region is unlikely to underpin potential differences in immune responses and, thus, in clinical outcomes among ethnic groups. It was notable, however, that within ethnic populations the capacity of individual HLA haplotypes to present viral peptides varied widely (data on HLA haplotypes examined and SARS-CoV-2 peptide presentation for the four broad ethnic populations is provided in **Supplemental Information**), with the top 5% of haplotypes predicted to present between 497 and 591 peptides from the entire viral proteome (such as DRB1*04:01~DRB4*01:01~DQA1*03:01~DQB1*03:01~DPA1*01:03~DPB1*04:01 which represents 1% of AFA haplotypes,

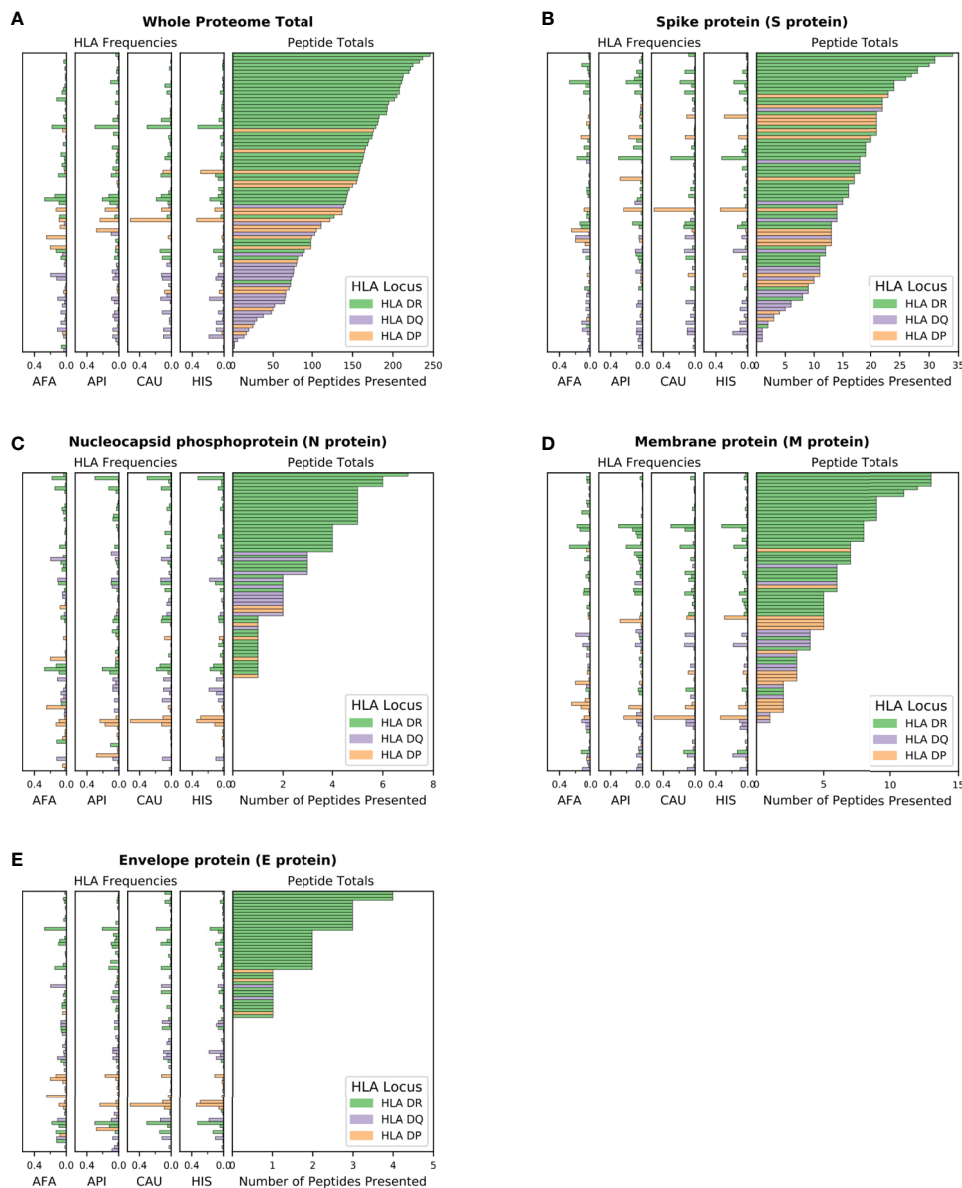


FIGURE 1 | SARS-CoV-2 viral proteome presentation at the molecular HLA class II level. The number of viral peptides presented by individual HLA is shown for **(A)** the entire SARS-CoV-2 proteome; **(B)** Spike protein (S protein); **(C)** Nucleocapsid protein (N protein); **(D)** Membrane protein (M protein); and **(E)** Envelope protein (E protein). The HLA frequency in four broad ethnic populations is shown (AFA, African Americans; API, Asian and Pacific Islanders; CAU, Caucasians; HIS, Hispanics). Bars are coloured according to HLA locus, as shown. HLA were included in the figure if they had a frequency of $\geq 1\%$ in any haplotype distribution from the four broad population groups.

0.14% of API haplotypes, 3.9% of CAU haplotypes and 0.9% of HIS haplotypes, predicted to present 506 SARS-CoV-2 peptides) as opposed to 5% of haplotypes at the opposite end of the spectrum predicted to present approximately 316 viral peptides or fewer (e.g. DRB1*03:01~DRB3*01:01~DQA1*05:01~DQB1*02:01~DPA1*02:01~DPB1*01:01 which represents 0.25% of AFA haplotypes, 0.02% of API haplotypes, 3.5% of CAU haplotypes and 1.4% of HIS haplotypes, predicted to present 258 peptides). This observation suggests that individual capacity to mount CD4⁺ T-cell immune responses against SARS-CoV-2 is

not uniform and is likely dependent on HLA phenotype. Similar inter-individual variability was noted for peptide presentation derived from structural and from non-structural proteins as well as for distinct SARS-CoV-2 proteins examined (**Figures 2B–E** and **S3**). Recent experimental studies suggested that up to 70% of CD4⁺ T-cell responses against SARS-CoV-2 target the Spike, Membrane and Nucleocapsid antigens (10); our analysis showed the predicted CD4⁺ T-cell response to these structural proteins is highly variable at the haplotype level (**Figures 2B–D**). In agreement with the observed immunogenicity of Spike protein

in experimental studies (28–31), relatively high numbers of Spike-derived peptides were predicted to be recognised both within and across ethnic populations (**Figure 2B**). In contrast, our analysis suggests that on average 13.9% of HLA haplotypes within each population have low capacity (≤ 2 peptides) to present Nucleocapsid-derived peptides (e.g. DRB1*03:02~DRB3*01:01~DQA1*04:01~DQB1*04:02~DPA1*02:02~DPB1*01:01 which presents one peptide, and accounts for 4.98% of haplotypes found in AFA populations and 0.44% in HIS populations, although it is rare in API and CAU populations; **Figure 2C**). This inter-individual variability may, in part, account for the heterogeneity in the presence and magnitude of CD4⁺ T-cell and antibody responses against the Nucleocapsid protein noted in recent COVID-19 studies (10, 32, 33). Among non-structural viral proteins, our analysis suggested NSP3, NSP4 and NSP12 as the most immunogenic, in part

reflecting their size, in each population (**Figure S3**). The above noted similarity in overall capacity for SARS-CoV-2 peptide presentation among different populations was also observed at different (more stringent) thresholds for HLA-peptide binding, albeit with even higher individual variability within ethnic populations (500nM binding affinity and $\leq 0.5\%$ percentage rank or ≤ 50 nM binding affinity; **Figure S4**).

To further explore the consequences of the above observations at the individual level, and given that every individual expresses two HLA haplotypes, we generated 10 replicates each of random genotype datasets encompassing 1,000, 5,000 and 10,000 simulated individuals for each population, as described in the methods. Populations encompassing 10,000 individuals achieved >95% cumulative HLA haplotype coverage in every population. As shown in **Figure 3**, this analysis confirmed equivalent HLA class II

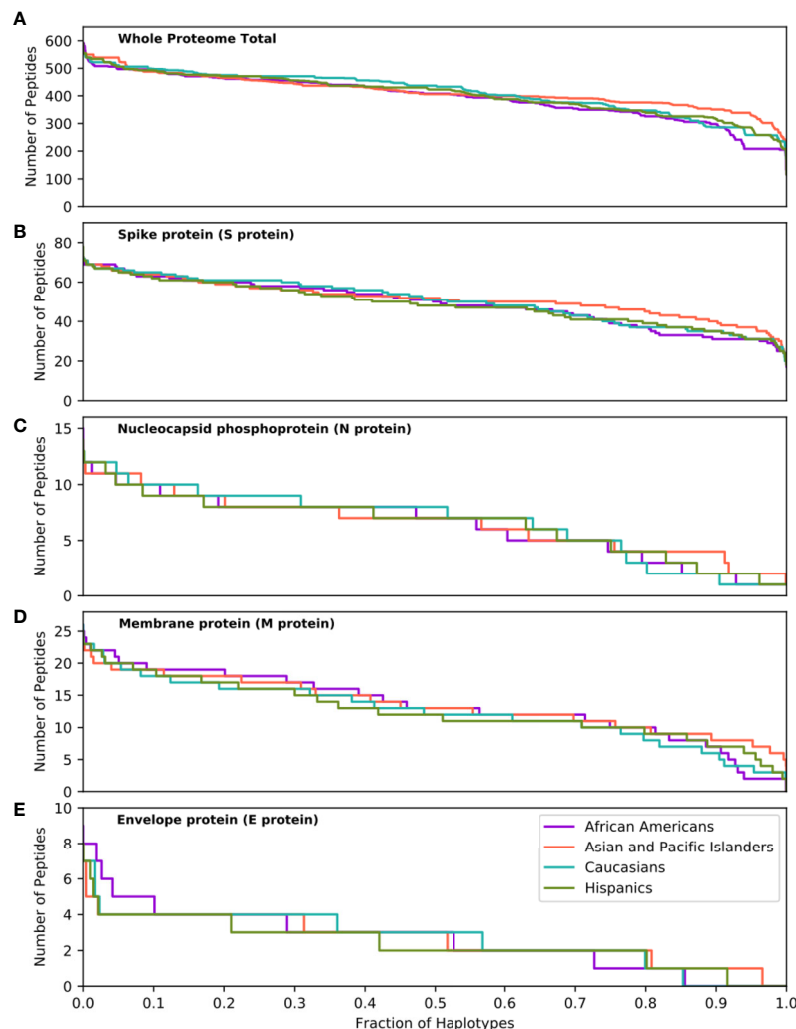


FIGURE 2 | SARS-CoV-2 derived peptide presentation at the HLA haplotype level. Panels depict SARS-CoV-2 peptide presentation by HLA class II haplotypes representing 99% of total haplotypes within four broad population groups (African Americans, Asian Pacific Islanders, Caucasians and Hispanics). **(A)** Whole Proteome. **(B)** Spike protein (S protein). **(C)** Nucleocapsid protein (N protein). **(D)** Membrane protein (M protein). **(E)** Envelope protein (E protein). The width of each step in the curves is proportional to the relative frequency of a specific HLA haplotype.

presentation of SARS-CoV-2 peptides across all four broad ethnic populations, both at the entire viral proteome level and for individual proteins (Table S4). As noted for the HLA haplotype analysis, there was wide inter-individual variability in predicted potential for CD4⁺ T-cell immune responses, according to HLA genotype. We noted significant, but variable, capacity for T-cell reactivity against the entire Spike glycoprotein across individuals, whereas reactivity against the Nucleocapsid

protein was predicted to be weaker for 10% of individuals in each population, on average (range 7.8-13.3%, as defined by HLA presentation of less than 5 nucleocapsid peptides). These observations were confirmed (but, again, inter-individual variability was higher) using more stringent thresholds for defining HLA class II peptide presentation (data not shown). In further analysis, we considered immune reactivity against the Receptor Binding Domain (RBD) of Spike glycoprotein, as it

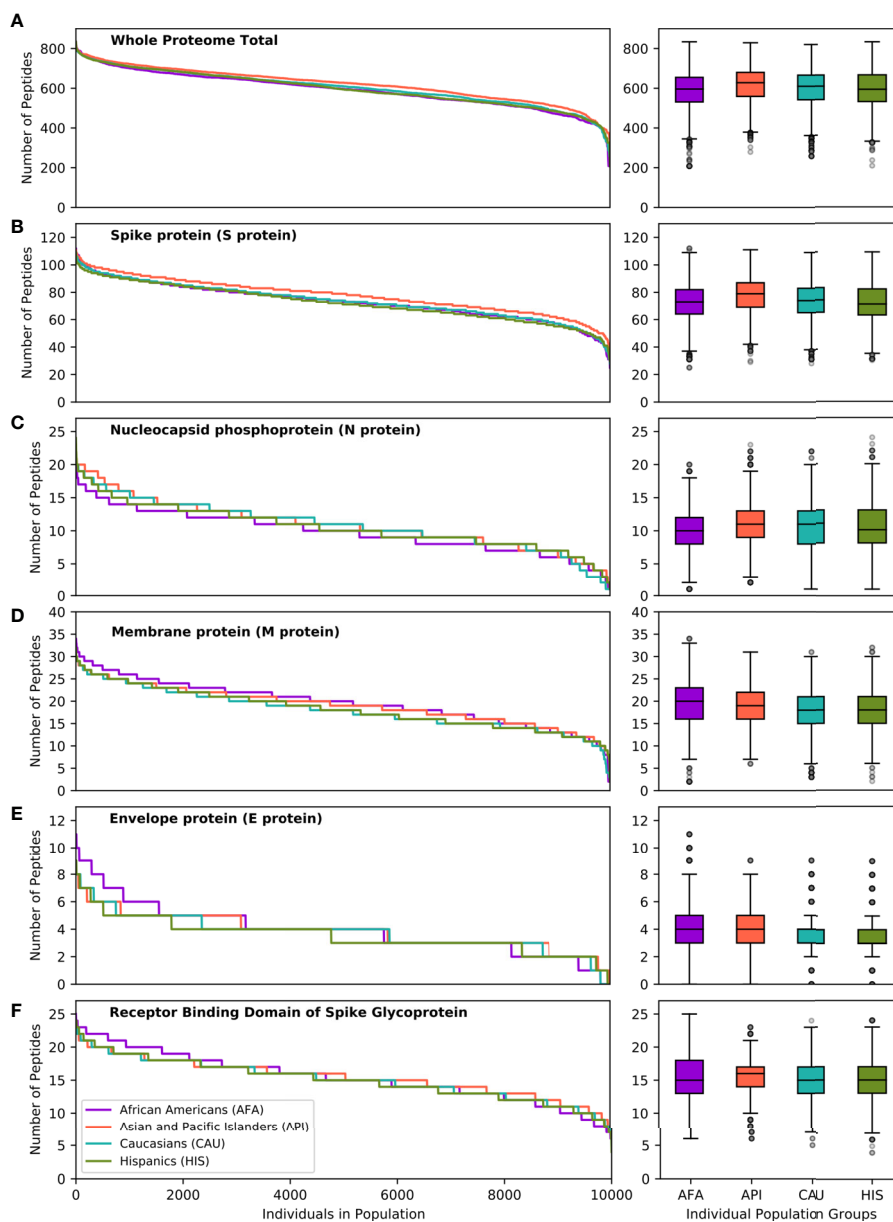


FIGURE 3 | SARS-CoV-2 derived peptide presentation at the HLA genotype level (broad population groups). Panels depict the number of SARS-CoV-2 peptides presented by individual HLA class II genotypes in simulated populations of 10,000 individuals for four broad population groups (African Americans, Asian Pacific Islanders, Caucasians and Hispanics). **(A)** Whole Proteome. **(B)** Spike protein (S protein). **(C)** Nucleocapsid protein (N protein). **(D)** Membrane protein (M protein). **(E)** Envelope protein (E protein). **(F)** Receptor Binding Domain of Spike protein. The width of each step in the curves is proportional to the relative frequency of a specific HLA genotype. The boxplot charts depict the median, interquartile range (box) and range (whiskers - excluding outliers) for the number of viral peptides presented at the population level for each of the above ethnic groups.

represents a proposed target of coronavirus subunit vaccines currently in clinical trials (34–36). Although, overall, there was no significant difference in predicted CD4⁺ T-cell reactivity at the population level, there were notable differences at the individual level (both based on HLA haplotype and on genotype analyses) with wide variation in predicted RBD specific peptide presentation (**Figure 3F**). The above analyses were consistent irrespective of the size of the population sampled and among the 10 replicates at each population size (Kruskal-Wallis test p-value >0.05 for peptide comparisons at the population level).

We next examined HLA class II presentation of SARS-CoV-2 peptides for a further 21 detailed population subgroups, as described above (**Figure S5** and **Table S4**). Overall, analysis of the entire viral proteome identified similar capacity for HLA class II viral peptide presentation across population subgroups. This was also the case for structural proteins, including Spike and RBD, mirroring the findings above for the broad population groups. Although we did not identify SARS-CoV-2 vulnerability of particular populations at the HLA level, again we observed inter-individual variation in predicted cellular immunity within ethnic groups. This was reflected in the range of predicted viral peptide presentation within simulated populations of 10,000 individuals including 24-112 for Spike, 4-25 for RBD, 1-24 for Nucleocapsid and 208-854 for the entire proteome (**Table S4**).

Immunogenicity Maps of SARS-CoV-2 Proteome at the Population Level

The effectiveness of peptide and subunit vaccine formulations against SARS-CoV-2 depends on robust presentation by individual HLA class II molecules and, therefore, investigated vaccines should account for linkage disequilibrium and HLA haplotype frequencies in different ethnic populations to achieve universal coverage. Capitalising on extensive HLA haplotype frequency data from the NMDP/Be The Match registry, we generated maps of SARS-CoV-2 immunogenicity for the entire viral proteome. This analysis showed that each viral protein contains immunogenic peptide segments with variable degree of population coverage which is, on the whole, similar across different populations for a given protein region, as well as peptide segments of variable length that are non-immunogenic in any population (**Figure 4**). Similar observations were made for coronavirus subunit components that are being investigated as potential vaccines with several immunogenic peptides predicted to achieve universal coverage across population groups (**Figure 4**). This was particularly evident for the Spike protein underlying its inherent immunogenicity and its potential as vaccination target. **Table S5**, depicts SARS-CoV-2 immunogenic peptide segments predicted to cover over 90% of HLA genotypic variation in every broad ethnic group examined.

DISCUSSION

Recognition of SARS-CoV-2 peptides in the context of HLA class II molecules is essential for CD4⁺ T-cell activation and proliferation which, in turn, orchestrate the development of

effector cellular (CD8⁺ T-cell) and humoral adaptive immune responses after viral infection and after vaccination. In this computational study, we investigated the role of HLA on SARS-CoV-2 immunogenicity at the individual and at the population level, considering population-specific HLA allele, haplotype, and genotype frequencies. We accounted for genetic polymorphism and for HLA linkage disequilibrium in twenty five ethnic populations by capitalising on, to our knowledge, the most extensive HLA haplotype frequency information to date to predict SARS-CoV-2 specific CD4⁺ T-cell epitopes covering the entire viral proteome. We find that the overall capacity for anti SARS-CoV-2 cellular immunity according to HLA class II genotype is similar at the population level across all ethnic groups examined. However, we identify wide inter-individual variability in predicted CD4⁺ T-cell reactivity against every SARS-CoV-2 protein according to expressed HLA genotype. We predict robust immune reactivity against the SARS-CoV-2 Spike protein, the basis for the majority of current vaccination efforts, both at the population and at the individual level regardless of population origin. Finally, we provide comprehensive maps of SARS-CoV-2 proteome immunogenicity accounting for population coverage in ethnic groups.

Several recent studies have examined the cellular immune response to SARS-CoV-2 and revealed strong associations between the T-cell response and COVID-19 severity (29, 37). Although this relationship is complex to untangle when the peripheral T-cell repertoire is sampled during the acute phase, it is notable that SARS-CoV-2 specific CD4⁺ T-cells have been associated with lessened COVID-19 severity and that high frequency of Spike-specific CD4⁺ T-cell responses were observed in patients who had recovered from COVID-19 (10, 29, 38, 39). A coordinated and regulated response involving all branches of adaptive immunity (CD4⁺, CD8⁺ and antibody responses) is likely required to reduce COVID-19 severity, with the cellular response being key for both initiating the adaptive response and for controlling the acute infection (29). Even though neutralising antibody titres are not predictive of disease severity (29, 40), humoral responses are a key aspect of protective immunity after infection and critical for generating sterilising immunity after vaccination (34, 41). In this respect, current evidence suggests a strong association between the magnitude of Spike-specific CD4⁺ T-cells and neutralising antibody titres (10, 30, 38). Finally, the majority of recent literature on anti-SARS-CoV-2 immunity indicates there is a high degree of heterogeneity in the breadth and magnitude of both humoral and cellular responses to SARS-CoV-2 both at the individual patient level and in relation to specific viral proteins (10, 29, 42). Within this context, the observation that patients with severe COVID-19 had decreased diversity of T-cell responses suggested that recognition of multiple SARS-CoV-2 T-cell epitopes may be required for development of protective immunity after infection or after vaccination (43).

The above observations on anti-SARS-CoV-2 cellular immunity from experimental studies and the role of HLA in shaping the diversity of the T-cell repertoire (44, 45), place the

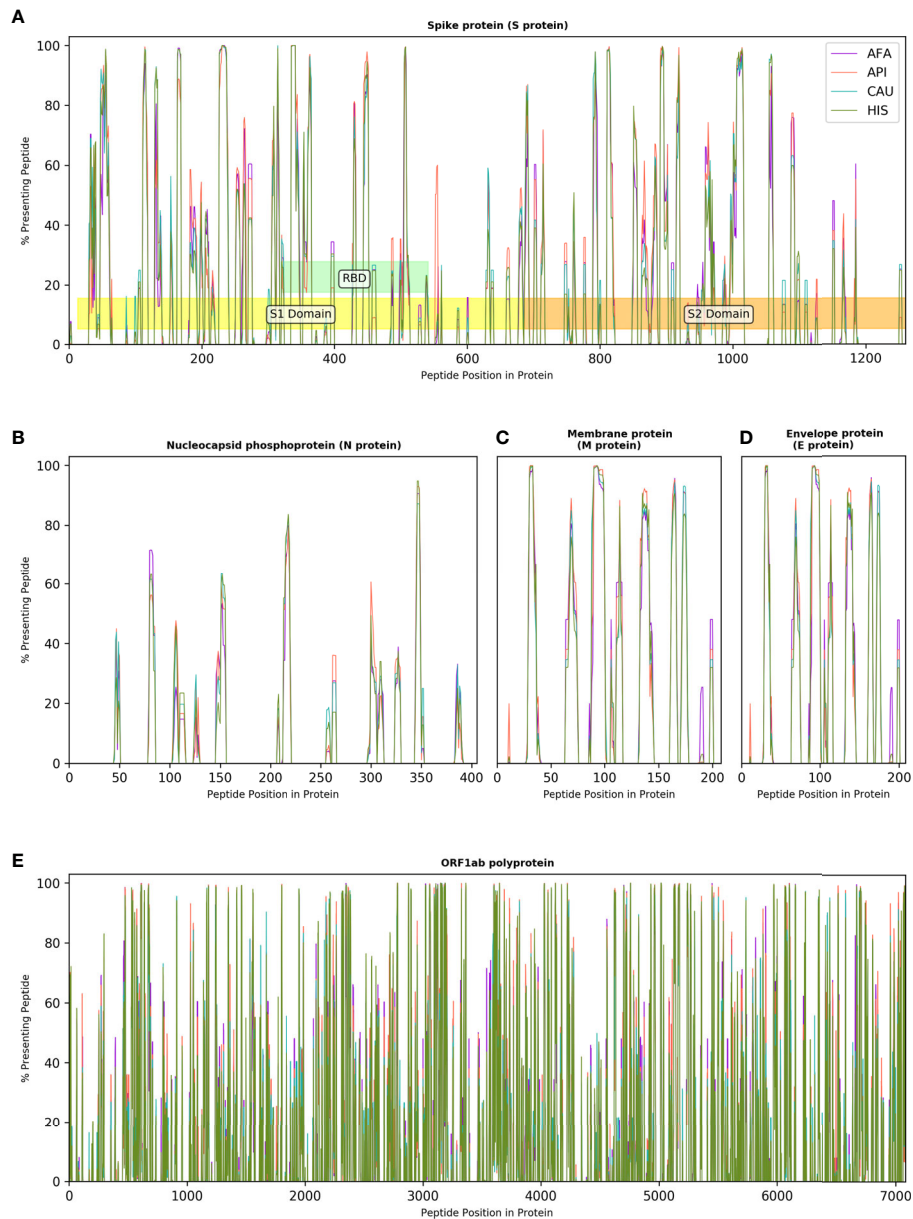


FIGURE 4 | Immunogenicity maps of SARS-CoV-2 proteome at the population level. The panels depict the percentage of HLA class II genotypes, within populations of 10,000 individuals (AFA, African Americans; API, Asian Pacific Islanders; CAU, Caucasians and HIS, Hispanics), that are predicted to present individual SARS-CoV-2 peptides from **(A)**. Spike protein (S protein), S1, S2 and Receptor Binding Domain (RBD) are also shown; **(B)**. Nucleocapsid protein (N protein); **(C)**. Membrane protein (M protein); **(D)**. Envelope protein (E protein); and **(E)**. Open Reading Frame (ORF) 1ab polyprotein.

findings of our study into context. With regards to susceptibility to COVID-19, we hypothesised that genetic variation at the HLA complex may account for observed differences in clinical outcomes between ethnic groups (2, 46, 47). We performed a comprehensive analysis of HLA haplotypes and genotypes covering 99% of genetic HLA variation within twenty five ethnic populations and showed that the predicted CD4⁺ T-cell response is overall remarkably similar at the population level both looking at the entire SARS-CoV-2 proteome and for individual

viral proteins. This observation is supported by more nuanced recent investigations which show equivalent COVID-19 clinical outcomes after adjustment for potential socioeconomic and clinical confounders (7, 8). We did, however, find significant inter-individual variability in predicted SARS-CoV-2 proteome immunogenicity according to HLA phenotype. This variability was more pronounced for particular viral proteins, such as nucleocapsid, membrane protein and envelope protein and less evident when the entire viral proteome was considered where,

even at the lower end of the spectrum, HLA haplotypes were predicted to present a significant number of CD4⁺ T-cell epitopes. Given the relevance of diversity and magnitude of cellular immunity against SARS-CoV-2, as discussed above, it is tempting to speculate that HLA phenotype might underpin some of the observed inter-individual variability in COVID-19 outcomes, along with more established clinical factors. This might depend on the relative contribution of SARS-CoV-2 proteins to the quality of the immune response. For example, Grifoni et al. (10) have shown that up to 70% of CD4⁺ T-cell responses against SARS-CoV-2 target the Spike, Membrane and Nucleocapsid antigens whereas significant reactivity was also noted against nsp3, nsp4 and ORF8. Despite significant differences between the highest and lowest peptide presenting HLA haplotypes, we noted substantial numbers of Spike-specific CD4⁺ T-cell peptides presented by the majority of HLA haplotypes in all ethnic groups examined; in comparison, inter-individual variability according to HLA haplotype was more pronounced for the remaining of the above, and other, viral proteins. Whether this observation might translate into differential clinical outcomes or levels of protective immunity according to expressed HLA type would need to be examined in large clinical studies that encompass cohorts representative of the HLA polymorphism within particular populations. Certainly, evidence supporting an important role of HLA class II in viral immunity has been previously reported (48, 49). The current SARS-CoV-2 pandemic represents a unique opportunity to address such fundamental questions which have recently started to be explored (50, 51).

It is also well established that individual response, including efficacy and relative antibody levels, and maintenance of immunity after vaccination varies markedly and this biological variation results from a combination of environmental (such as age, size, sex, comorbid status, ethnicity, and dose and route of vaccine administration) and genetic factors (12, 49). Non-responsiveness affects approximately 2-10% (and up to 20% following hepatitis B vaccination) of vaccinated healthy individuals (52, 53). HLA class II haplotype plays a central role in the presentation of vaccine epitopes and is a known genetic risk factor for primary vaccination failure (52, 54, 55). Given that the majority of current vaccination efforts are focused on generating immunity against the Spike protein of SARS-CoV-2, we calculated the number of Spike-specific CD4⁺ T-cell epitopes according to HLA genotype. Our analysis suggests that immune reactivity against Spike is likely to be robust both at the population level, including all 25 ethnic groups examined, and at the individual level. This finding is now supported experimentally by studies reporting high degree of seroconversion against Spike after natural infection and after vaccination (28, 30, 42, 56–61). Nevertheless, we noted inter-individual variation ranging from 112 peptides for the highest presenting HLA class II genotypes to 24 for the lowest. Whether such variation may affect the magnitude and diversity of protective immunity generated after vaccination requires further study but it is notable that variation in the degree of cellular immunity has been reported with a few vaccine formulations (57, 60, 62). To our knowledge, the relationship

between HLA genotype, number of vaccine-derived T-cell epitopes and vaccine responsiveness has not been systematically examined yet, although this is the focus of current research efforts in the context of COVID-19.

It is important to acknowledge the limitations of our study. We used a computational approach to predict SARS-CoV-2 peptides presented by HLA class II molecules, however, peptide presentation does not always lead to CD4⁺ T-cell activation; peptide recognition is complex and incompletely understood and is influenced by many factors, including relative expression of individual viral proteins (63). Nevertheless, NetMHCIIpan-4.0 is an established and validated algorithm for T-cell epitope prediction that has recently been updated resulting in improved performance (19). Recent computational studies investigating SARS-CoV-2 vaccine immunogenicity have based their approach for T-cell epitope selection exclusively on peptide-HLA binding affinity incorporating different thresholds (e.g. 500nM or 50nM) and identified population coverage gaps in predicted cellular immunity (13, 64, 65). This approach is affected by inherent bias of certain HLA molecules towards higher or lower mean predicted affinities; thus, we show that the 50nM binding affinity threshold, one of the most commonly used, is heavily biased towards HLA-DR as the main SARS-CoV-2 peptide presenting locus with the majority of HLA-DQ and -DP molecules showing no peptide binding. Accordingly, using a 50nM binding affinity threshold for defining peptide immunogenicity resulted in very wide inter-individual variability in predicted CD4⁺ T-cell reactivity against SARS-CoV-2 proteins (**Figure S4**). To overcome this limitation, we incorporated both binding affinity prediction and percentile rank, compared to a set of 100,000 random natural peptides, for epitope selection. The rank score normalizes prediction scores across different HLA molecules and enables interspecific HLA binding prediction comparisons; nevertheless, part of the difference in the peptide binding capacity of HLA molecules (highest for HLA-DR) in this study may reflect training of NetMHCIIpan-4.0 algorithm on peptide datasets that contain more information on HLA-DR restricted peptides. Notwithstanding that currently there is limited information on experimentally determined SARS-CoV-2 immunogenic T-cell epitopes and even less information on peptide HLA restriction, we have used available information in the published literature to validate our approach. We focused on prediction of strongly HLA binding peptides and demonstrated that our peptide selection threshold correctly identified the majority of experimentally determined immunogenic viral peptides in the largest published datasets (20, 23). We also aimed to minimise the rate of false positive peptides and showed, in a limited experimental SARS-CoV-2 peptide dataset with HLA restrictions published by Prachar et al, that our 500nM binding affinity and $\leq 2\%$ rank threshold for peptide selection, achieves high precision and specificity (1.0 for both). Also, we confirmed that our observations on SARS-CoV-2 immunogenicity, both in relation to the comparison of population level responses and in relation to inter-individual variability, remained valid when we used more stringent peptide selection criteria (500nM and $\leq 0.5\%$ rank) and the commonly

used 50nM binding affinity threshold. Finally, we used the NMDP/Be The Match registry to compute HLA population frequencies. In terms of population genetics, this study is limited in that the breadth of HLA diversity of global human populations is incompletely represented by the US population categories, and sampling depth varied widely among those categories included. However, the accuracy of US population HLA frequency estimates has been validated in multiple practical settings in transplantation, and HLA frequencies from many global population datasets from other stem cell registries often have high similarity with US population estimates (66–68).

In conclusion, we present a rigorous immune-informatics approach to evaluate the potential for cellular immunity against SARS-CoV-2 at the population and at the individual level capitalising on, to our knowledge, the most comprehensive assessment of HLA genetic variation to date. Our findings provide important insight on the potential role of HLA polymorphism on development of protective immunity after SARS-CoV-2 infection and after vaccination and a firm basis for further experimental studies in this field.

DATA AVAILABILITY STATEMENT

The datasets generated and analysed for this study has been deposited in Mendeley Data: HLA haplotype population frequency data - US unrelated stem cell donor registry National Marrow Donor Program (NMDP) / Be The Match (69) and Data on HLA haplotypes examined and SARS-CoV-2 peptide presentation for the four broad ethnic populations (70).

AUTHOR CONTRIBUTIONS

HC, VK, AL and LG contributed to definition of the central question and solutions. HC developed and implemented the peptide-HLA prediction models, the HLA, haplotype and population analysis and the algorithm validation with advice and supervision from AL and VK. LG performed the HLA

population frequency computation, calculated the Haplotype frequency distributions, generated simulated populations and performed the statistical analysis on simulated populations. VK conceived and supervised the work and wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

The research presented in this manuscript has received funding from the NIHR Blood and Transplant Research Unit in Organ Donation and Transplantation at the University of Cambridge, from the NIHR Cambridge Biomedical Research Centre and from an NIHR Fellowship (PDF-2016-09-065, VK). We gratefully acknowledge funding from an MRC Clinical Research Training Fellowship to HC (MR/S006745/1). VL acknowledges funding as a P.I. Terasaki Scholar. AL acknowledges funding by the Member States of the European Molecular Biology Laboratory (EMBL).

ACKNOWLEDGMENTS

We thank the US Registry National Marrow Donor Program/Be The Match and the 38+ million stem cell registry volunteers worldwide. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health Research, the Department of Health, or National Health Service Blood and Transplant.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2021.669357/full#supplementary-material>

REFERENCES

1. WHO_COVID-19_Dashboard. Available at: <https://covid19.who.int/>.
2. Williamson EJ, Walker AJ, Bhaskaran K, Bacon S, Bates C, Morton CE, et al. Factors Associated With COVID-19-Related Death Using Opensafely. *Nature* (2020) 584:430. doi: 10.1038/s41586-020-2521-4
3. Grasselli G, Zangrillo A, Zanella A, Antonelli M, Cabrini L, Castelli A, et al. Baseline Characteristics and Outcomes of 1591 Patients Infected With SARS-CoV-2 Admitted to ICUs of the Lombardy Region, Italy. *JAMA* (2020) 323:1574. doi: 10.1001/jama.2020.5394
4. Chow N, Fleming-Dutra K, Gierke R, Hall A, Hughes M, Pilishvili T, et al. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. *MMWR Morbidity Mortality Weekly Rep* (2020) 69:382. doi: 10.15585/mmwr.mm6913e2
5. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With Covid-19 in the New York City Area. *JAMA* (2020) 323:2052. doi: 10.1001/jama.2020.6775
6. Hsu HE, Ashe EM, Silverstein M, Hofman M, Lange SJ, Razzaghi H, et al. Race/Ethnicity, Underlying Medical Conditions, Homelessness, and Hospitalization Status of Adult Patients With COVID-19 at an Urban Safety-Net Medical Center - Boston, Massachusetts, 2020. *MMWR Morbidity Mortality Weekly Rep* (2020) 69:864. doi: 10.15585/mmwr.mm6927a3
7. Price-Haywood EG, Burton J, Fort D, Seoane L. Hospitalization and Mortality Among Black Patients and White Patients With Covid-19. *N Engl J Med* (2020) 382:2534. doi: 10.1056/NEJMsa2011686
8. Yehia BR, Winegar A, Fogel R, Fakhri M, Ottenbacher A, Jessor C, et al. Association of Race With Mortality Among Patients Hospitalized With Coronavirus Disease 2019 (Covid-19) at 92 US Hospitals. *JAMA Network Open* (2020) 3:e2018039. doi: 10.1001/jamanetworkopen.2020.18039
9. Zhang Q, Bastard P, Liu Z, Le Pen J, Moncada-Velez M, Chen J, et al. Inborn Errors of Type I IFN Immunity in Patients With Life-Threatening COVID-19. *Science* (2020) 370:eabd4570. doi: 10.1126/science.abd4570
10. Grifoni A, Weiskopf D, Ramirez SI, Mateus J, Dan JM, Rydzynski Moderbacher C, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans With COVID-19 Disease and Unexposed Individuals. *Cell* (2020) 181:1489. doi: 10.1016/j.cell.2020.05.015

11. Staines HM, Kirwan DE, Clark DJ, Adams ER, Augustin Y, Byrne RL, et al. Igg Seroconversion and Pathophysiology in Severe Acute Respiratory Syndrome Coronavirus 2 Infection. *Emerging Infect Dis* (2021) 27:85. doi: 10.3201/eid2701.203074
12. Mentzer AJ, O'Connor D, Pollard AJ, Hill AVS. Searching for the Human Genetic Factors Standing in the Way of Universally Effective Vaccines. *Philos Trans R Soc Lond B Biol Sci* (2015) 370:20140341. doi: 10.1098/rstb.2014.0341
13. Liu G, Carter B, Bricken T, Jain S, Viard M, Carrington M, et al. Computationally Optimized SARS-Cov-2 MHC Class I and II Vaccine Formulations Predicted to Target Human Haplotype Distributions. *Cell Syst* (2020) 11:131. doi: 10.1016/j.cels.2020.06.009
14. Consortium TU. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res* (2018) 47:D506. doi: 10.1093/nar/gky1049
15. Gragert L, Madbouly A, Freeman J, Maiers M. Six-Locus High Resolution HLA Haplotype Frequencies Derived From Mixed-Resolution DNA Typing for the Entire US Donor Registry. *Hum Immunol* (2013) 74:1313. doi: 10.1016/j.humimm.2013.06.025
16. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M, et al. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved Predictions of MHC Antigen Presentation by Concurrent Motif Deconvolution and Integration of MS MHC Eluted Ligand Data. *Nucleic Acids Res* (2020) 48:W449. doi: 10.1093/nar/gkaa379
17. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-Learn: Machine Learning in Python. *J Mach Learn Res* (2011) 12:2825. doi: 10.5555/1953048.2078195
18. Wang C, Liu Z, Chen Z, Huang X, Xu M, He T, et al. The Establishment of Reference Sequence for SARS-CoV-2 and Variation Analysis. *J Med Virol* (2020) 92:667. doi: 10.1002/jmv.25762
19. Reynisson B, Barra C, Kaabinejadian S, Hildebrand WH, Peters B, Nielsen M, et al. Improved Prediction of MHC II Antigen Presentation Through Integration and Motif Deconvolution of Mass Spectrometry MHC Eluted Ligand Data. *J Proteome Res* (2020) 19:2304. doi: 10.1021/acs.jproteome.9b00874
20. Snyder TM, Gittelman RM, Klinger M, May DH, Osborne EJ, Taniguchi R, et al. Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. *medRxiv Preprint Server Health Sci* (2020). doi: 10.1101/2020.07.31.20165647
21. Liu G, Carter B, Gifford DK. Predicted Cellular Immunity Population Coverage Gaps for SARS-CoV-2 Subunit Vaccines and Their Augmentation by Compact Peptide Sets. *Cell Syst* (2021) 102-7. doi: 10.1016/j.cels.2020.11.010
22. Nolan S, Vignali M, Klinger M, Dines JN, Kaplan IM, Svejnoha E, et al. A Large-Scale Database Of T-Cell Receptor Beta (Tcr β) Sequences And Binding Associations From Natural And Synthetic Exposure To Sars-Cov-2. (2020). doi: 10.21203/rs.3.rs-51964/v1
23. Mateus J, Grifoni A, Tarke A, Sidney J, Ramirez SI, Dan JM, et al. Selective and Cross-Reactive SARS-Cov-2 T Cell Epitopes in Unexposed Humans. *Science* (2020) 370:89. doi: 10.1126/science.abd3871
24. Le Bert N, Tan AT, Kunasegaran K, Tham CYL, Hafezi M, Chia A, et al. SARS-Cov-2-Specific T Cell Immunity in Cases of COVID-19 and SARS, and Uninfected Controls. *Nature* (2020) 584:457. doi: 10.1038/s41586-020-2550-z
25. Keller MD, Harris KM, Jensen-Wachspress MA, Kankate V, Lang H, Lazarski CA, et al. SARS-Cov-2-Specific T Cells Are Rapidly Expanded for Therapeutic Use and Target Conserved Regions of the Membrane Protein. *Blood* (2020) 136:2905. doi: 10.1182/blood.2020008488
26. Prachar M, Justesen S, Steen-Jensen DB, Thorgrimsen S, Jurgons E, Winther O, et al. Identification and Validation of 174 COVID-19 Vaccine Candidate Epitopes Reveals Low Performance of Common Epitope Prediction Tools. *Sci Rep* (2020) 10:20465. doi: 10.1038/s41598-020-77466-4
27. Tarke A, Sidney J, Kidd CK, Dan JM, Ramirez SI, Yu ED, et al. Comprehensive Analysis of T Cell Immunodominance and Immunoprevalence of SARS-CoV-2 Epitopes in COVID-19 cases. *Cell Rep Med* (2021) 2(2):100204. doi: 10.1016/j.xcrm.2021.100204
28. Wajnberg A, Mansour M, Leven E, Bouvier NM, Patel G, Firpo-Betancourt A, et al. Humoral Response and PCR Positivity in Patients With COVID-19 in the New York City Region, USA: An Observational Study. *Lancet Microbe* (2020) 1:e283. doi: 10.1016/S2666-5247(20)30120-8
29. Rydzynski Moderbacher C, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, et al. Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute Covid-19 and Associations With Age and Disease Severity. *Cell* (2020) 183:996. doi: 10.1016/j.cell.2020.09.038
30. Juno JA, Tan H-X, Lee WS, Reynaldi A, Kelly HG, Wragg K, et al. Humoral and Circulating Follicular Helper T Cell Responses in Recovered Patients With COVID-19. *Nat Med* (2020) 26:1428. doi: 10.1038/s41591-020-0995-0
31. Iyer AS, Jones FK, Nodoushani A, Kelly M, Becker M, Slater D, et al. Persistence and Decay of Human Antibody Responses to the Receptor Binding Domain of SARS-CoV-2 Spike Protein in COVID-19 Patients. *Sci Immunol* (2020) 5:eabe0367. doi: 10.1126/sciimmunol.abe0367
32. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Pérez-Olmeda M, Yotti R, et al. Prevalence of SARS-CoV-2 in Spain (Ene-COVID): A Nationwide, Population-Based Seroepidemiological Study. *Lancet* (2020) 396:535. doi: 10.1016/S0140-6736(20)32266-2
33. Dan JM, Mateus J, Kato Y, Hastie KM, Dawen E, Faliti CE, et al. Immunological Memory to SARS-CoV-2 Assessed for Up to 8 Months After Infection. *Science* (2021) 371:eabf4063. doi: 10.1126/science.abf4063
34. Yu J, Tostanoski LH, Peter L, Mercado NB, McMahan K, Mahrokhian SH, et al. DNA Vaccine Protection Against SARS-CoV-2 in Rhesus Macaques. *Science* (2020) 369:806. doi: 10.1126/science.abc6284
35. Dai L, Zheng T, Xu K, Han Y, Xu L, Huang E, et al. A Universal Design of Betacoronavirus Vaccines against COVID-19, MERS, and SARS. *Cell* (2020) 182:722. doi: 10.1016/j.cell.2020.06.035
36. Dong Y, Dai T, Wei Y, Zhang L, Zheng M, Zhou F, et al. A Systematic Review of SARS-CoV-2 Vaccine Candidates. *Signal Transduction Targeted Ther* (2020) 5:237. doi: 10.1038/s41392-020-00352-y
37. Vabret N, Britton GJ, Gruber C, Hegde S, Kim J, Kuksin M, et al. Immunology of COVID-19: Current State of The Science. *Immunity* (2020) 52:910. doi: 10.1016/j.immuni.2020.05.002
38. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, et al. Broad and Strong Memory CD4+ and CD8+ T Cells Induced by SARS-CoV-2 in UK Convalescent Individuals Following COVID-19. *Nat Immunol* (2020) 21:1336. doi: 10.1038/s41590-020-0782-6
39. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Strålin K, Gorin JB, Olsson A, et al. Robust T Cell Immunity in Convalescent Individuals With Asymptomatic or Mild Covid-19. *Cell* (2020) 183:158. doi: 10.1016/j.cell.2020.08.017
40. Huang M, Lu QB, Zhao H, Zhang Y, Sui Z, Fang L, et al. Temporal Antibody Responses to SARS-CoV-2 in Patients of Coronavirus Disease 2019. *Cell Discovery* (2020) 6:64. doi: 10.1038/s41421-020-00209-2
41. Corbett KS, Flynn B, Foulds KE, Francica JR, Boyoglu-Barnum S, Werner AP, et al. Evaluation of the mRNA-1273 Vaccine Against SARS-CoV-2 in Nonhuman Primates. *New Engl J Med* (2020) 383:1544. doi: 10.1056/NEJMoa2024671
42. Dan JM, Mateus J, Kato Y, Hastie KM, Faliti CE, Ramirez SI, et al. Immunological Memory to SARS-CoV-2 Assessed for Greater Than Six Months After Infection. *bioRxiv* (2020). doi: 10.1101/2020.11.15.383323
43. Nelde A, Bilich T, Heitmann JS, Maringer Y, Salih HR, Roerden M, et al. SARS-CoV-2-derived Peptides Define Heterologous and COVID-19-Induced T Cell Recognition. *Nat Immunol* (2021) 22:74. doi: 10.1038/s41590-020-00808-x
44. Goldrath AW, Bevan MJ. Selecting and Maintaining a Diverse T-Cell Repertoire. *Nature* (1999) 402:255. doi: 10.1038/46218
45. Messaoudi I, Guevara Patiño JA, Dyall R, LeMaout J, Nikolich-Zugich J, et al. Direct Link Between Mhc Polymorphism, T Cell Avidity, and Diversity in Immune Defense. *Science* (2002) 298:1797. doi: 10.1126/science.1076064
46. Lassale C, Gaye B, Hamer M, Gale CR, Batty GD, et al. Ethnic Disparities in Hospitalisation for COVID-19 in England: The Role of Socioeconomic Factors, Mental Health, and Inflammatory and Pro-Inflammatory Factors in a Community-Based Cohort Study. *Brain Behav Immun* (2020) 88:44. doi: 10.1016/j.bbi.2020.05.074
47. Aldridge RW, Lewer D, Katikireddi SV, Mathur R, Pathak N, Burns R, et al. Black, Asian and Minority Ethnic Groups in England are at Increased Risk of Death From COVID-19: Indirect Standardisation of NHS Mortality Data. *Wellcome Open Res* (2020) 5:88. doi: 10.12688/wellcomeopenres.15922.1
48. Guo X, Zhang Y, Li J, Ma J, Wei Z, Tan W, et al. Strong Influence of Human Leukocyte Antigen (HLA)-DP Gene Variants on Development of Persistent Chronic Hepatitis B Virus Carriers in the Han Chinese Population. *Hepatology* (2011) 53:422. doi: 10.1002/hep.24048
49. Blackwell JM, Jamieson SE, Burgner D. HLA and Infectious Diseases. *Clin Microbiol Rev* (2009) 22:370. doi: 10.1128/CMR.00048-08

50. Pisanti S, Deelen J, Gallina AM, Caputo M, Citro M, Abate M, et al. Correlation of the Two Most Frequent HLA Haplotypes in the Italian Population to the Differential Regional Incidence of Covid-19. *J Trans Med* (2020) 18:352. doi: 10.1186/s12967-020-02515-5
51. Langton DJ, Bourke SC, Lie BA, Reiff G, Natu S, Darlay R, et al. The Influence Of Hla Genotype On Susceptibility To, And Severity Of, Covid-19 Infection. *medRxiv* (2021). doi: 10.1101/2020.12.31.20249081
52. Wiedermann U, Garner-Spitzer E, Wagner A. Primary Vaccine Failure to Routine Vaccines: Why and What to do? *Hum Vaccines Immunotherapeutics* (2016) 12:239. doi: 10.1080/21645515.2015.1093263
53. Kimman TG, Vandebriel RJ, Hoebee B. Genetic Variation in the Response to Vaccination. *Community Genet* (2007) 10:201. doi: 10.1159/000106559
54. Gelder CM, Lambkin R, Hart KW, Fleming D, Williams OM, Bunce M, et al. Associations Between Human Leukocyte Antigens and Nonresponsiveness to Influenza Vaccine. *J Infect Dis* (2002) 185:114. doi: 10.1086/338014
55. Godkin A, Davenport M, Hill AV. Molecular Analysis of HLA Class II Associations With Hepatitis B Virus Clearance and Vaccine Nonresponsiveness. *Hepatology* (2005) 41:1383. doi: 10.1002/hep.20716
56. Gregoriadis G. Liposomes In Therapeutic And Preventive Medicine: The Development Of The Drug-Carrier Concept. *Ann NY Acad Sci* (1978) 308:343. doi: 10.1111/j.1749-6632.1978.tb22034.x
57. Anderson EJ, Roupael NG, Widge AT, Jackson LA, Roberts PC, Makhene M, et al. Safety and Immunogenicity of SARS-CoV-2 mRNA-1273 Vaccine in Older Adults. *N Engl J Med* (2020) 383:2427. doi: 10.1056/NEJMoa2028436
58. Walsh EE, Frenck RW Jr., Falsey AR, Kitchin N, Absalon J, Gurtman A, et al. Safety and Immunogenicity of Two Rna-Based Covid-19 Vaccine Candidates. *N Engl J Med* (2020) 383:2439. doi: 10.1056/NEJMoa2027906
59. Zhu F-C, Li Y-H, Guan X-H, Hou LH, Wang WJ, Li JX, et al. Safety, Tolerability, and Immunogenicity of a Recombinant Adenovirus Type-5 Vectored COVID-19 Vaccine: A Dose-Escalation, Open-Label, Non-Randomised, First-in-Human Trial. *Lancet* (2020) 395:1845. doi: 10.1016/S0140-6736(20)31208-3
60. Ramasamy MN, Minassian AM, Ewer KJ, Flaxman AL, Folegatti PM, Owens DR, et al. Safety and Immunogenicity of ChAdOx1 nCoV-19 Vaccine Administered in a Prime-Boost Regimen in Young and Old Adults (COV002): A Single-Blind, Randomised, Controlled, Phase 2/3 Trial. *Lancet* (2021) 396:1979. doi: 10.1016/S0140-6736(20)32466-1
61. Logunov DY, Dolzhikova IV, Zubkova OV, Tukhvatulin AI, Shcheblyakov DV, Dzharullaeva AS, et al. Safety and Immunogenicity of an rAd26 and rAd5 Vector-Based Heterologous Prime-Boost COVID-19 Vaccine in Two Formulations: Two Open, non-Randomised Phase 1/2 Studies From Russia. *Lancet* (2020) 396:887. doi: 10.1016/S0140-6736(20)31866-3
62. Zhu F-C, Guan X-H, Li Y-H, Huang JY, Jiang T, Hou LH, et al. Immunogenicity and Safety of a Recombinant Adenovirus Type-5-Vectored COVID-19 Vaccine in Healthy Adults Aged 18 Years or Older: A Randomised, Double-Blind, Placebo-Controlled, Phase 2 Trial. *Lancet* (2020) 396:479. doi: 10.1016/S0140-6736(20)31605-6
63. Roche PA, Furuta K. The Ins and Outs of MHC Class II-Mediated Antigen Processing and Presentation. *Nat Rev Immunol* (2015) 15:203. doi: 10.1038/nri3818
64. Yarmarkovich M, Warrington JM, Farrel A, Maris JM. Identification of SARS-CoV-2 Vaccine Epitopes Predicted to Induce Long-Term Population-Scale Immunity. *Cell Rep Med* (2020) 1(3):100036. doi: 10.1016/j.xcrm.2020.100036
65. Behmard E, Soleymani B, Najafi A, Barzegari E. Immunoinformatic Design of a COVID-19 Subunit Vaccine Using Entire Structural Immunogenic Epitopes of SARS-Cov-2. *Sci Rep* (2020) 10:20864. doi: 10.1038/s41598-020-77547-4
66. Gragert L, Eapen M, Williams E, Freeman J, Spellman S, Baitty R, et al. HLA Match Likelihoods for Hematopoietic Stem-Cell Grafts in the U.S. Registry. *N Engl J Med* (2014) 371:339. doi: 10.1056/NEJMsa1311707
67. Simanovsky AL, Madbouly A, Halagan M, Maiers M, Louzoun Y, et al. Single Haplotype Admixture Models Using Large Scale HLA Genotype Frequencies to Reproduce Human Admixture. *Immunogenetics* (2019) 71:589. doi: 10.1007/s00251-019-01144-7
68. Madbouly A, Gragert L, Freeman J, Leahy N, Gourraud PA, Hollenbach JA, et al. Validation of Statistical Imputation of Allele-Level Multilocus Phased Genotypes From Ambiguous HLA Assignments. *Tissue Antigens* (2014) 84:285. doi: 10.1111/tan.12390
69. Available at: <https://data.mendeley.com/datasets/545r9cggzf/1>.
70. Available at: <https://data.mendeley.com/datasets/bjbw35cg35/1>.

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

Copyright © 2021 Copley, Gragert, Leach and Kosmoliaptis. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.