

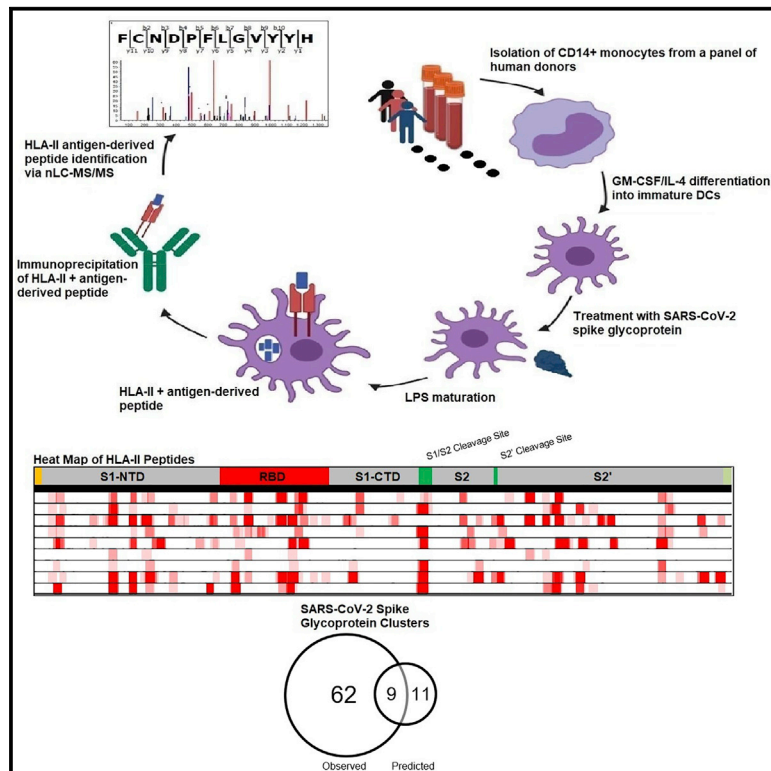


Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

The Human Leukocyte Antigen Class II Immunopeptidome of the SARS-CoV-2 Spike Glycoprotein

Graphical Abstract



Authors

Michael D. Knierman, Megan B. Lannan, Laura J. Spindler, Carl L. McMillian, Robert J. Konrad, Robert W. Siegel

Correspondence

siegel_robert@lilly.com

In Brief

Knierman et al. pulse dendritic cells derived from healthy human donors with the SARS-CoV-2 spike glycoprotein to determine the precise sequences presented for T cell surveillance. Regions with promiscuous presentation are identified, with poor correlation to predicted epitopes. One region with promiscuous presentation is conserved with the SARS-CoV spike glycoprotein sequence.

Highlights

- Dendritic cells display peptides spanning the entire SARS-CoV-2 spike protein
- HLA-II peptides from 11 regions are presented by a majority of the analyzed donors
- One region with promiscuous HLA-II presentation is conserved with SARS-CoV
- The correlation of presented to predicted peptides is low



Article

The Human Leukocyte Antigen Class II Immunopeptidome of the SARS-CoV-2 Spike Glycoprotein

Michael D. Knierman,¹ Megan B. Lannan,¹ Laura J. Spindler,¹ Carl L. McMillian,¹ Robert J. Konrad,¹ and Robert W. Siegel^{1,2,*}

¹Lilly Research Laboratories, Eli Lilly and Company, Indianapolis, IN 46285, USA

²Lead Contact

*Correspondence: siegel_robert@lilly.com

<https://doi.org/10.1016/j.celrep.2020.108454>

SUMMARY

Precise elucidation of the antigen sequences for T cell immunosurveillance greatly enhances our ability to understand and modulate humoral responses to viral infection or active immunization. Mass spectrometry is used to identify 526 unique sequences from the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) spike glycoprotein extracellular domain in a complex with human leukocyte antigen class II molecules on antigen-presenting cells from a panel of healthy donors selected to represent a majority of allele usage from this highly polymorphic molecule. The identified sequences span the entire spike protein, and several sequences are isolated from a majority of the sampled donors, indicating promiscuous binding. Importantly, many peptides derived from the receptor binding domain used for cell entry are identified. This work represents a precise and comprehensive immunopeptidomic investigation with the SARS-CoV-2 spike glycoprotein and allows detailed analysis of features that may aid vaccine development to end the current coronavirus disease 2019 (COVID-19) pandemic.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a positive-sense single-strand RNA virus that is a novel member of the genus *Betacoronavirus*, family Coronaviridae, responsible for coronavirus disease 2019 (COVID-19), which emerged in China in late 2019 and became a global pandemic by March 2020. At the time of manuscript preparation, over 34,000,000 cases with more than 1,000,000 fatalities have been reported worldwide (<https://coronavirus.jhu.edu/>). There are four additional members of the *Betacoronavirus* genus: two (human coronavirus [HCoV]-HKU1 and HCoV-OC43) that cause mild respiratory symptoms associated with the common cold and two (SARS-CoV and Middle East respiratory syndrome [MERS]-CoV) that can cause fatal respiratory tract infections. The SARS-CoV-2 genomic sequence shares 79.6% identity with SARS-CoV and 96.2% identity with SARS-related coronavirus RatG13, isolated from bats, supporting a zoonotic origin (Zhou et al., 2020a). At present, no therapeutic interventions to treat or prevent COVID-19 have been approved for use (<https://covid-19tracker.milkeninstitute.org/>).

Active immunization is an area of intense research, with over 100 programs under development, and successful implementation would greatly aid in ending the current pandemic (<http://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines>). Immunization strategies include use of live attenuated viruses, inactivated

viruses, non-replicating viral vectors, protein subunits, and nucleic acid-based approaches. At the time of manuscript preparation, at least 43 candidate vaccines have progressed into clinical evaluation. Effective immunization would prevent virus entry into cells. The surface spike glycoprotein is the defining feature of all coronaviruses and is critical for internalization by engaging the host receptor and mediation of virus-host membrane fusion (Cavanagh, 1995). The SARS-CoV-2 spike protein, as with SARS-CoV, interacts with human angiotensin-converting enzyme 2 (ACE2) (Zhou et al., 2020b; Walls et al., 2020; Li et al., 2020). The transmembrane glycoprotein is comprised of two functional subunits and forms homotrimers on the viral cell surface. A defined receptor binding domain (RBD) within the S1 subunit is responsible for binding to ACE2 (Shang et al., 2020; Li et al., 2020). The S2 subunit in all coronaviruses contains 2-heptad repeat segments that form a coiled-coil structure, and membrane fusion occurs after proteolytic processing and conformational rearrangement (Liu et al., 2004; Millet and Whittaker, 2015). A 4-amino-acid polybasic furin cleavage site insertion between the S1 and S2 subunits is a unique feature of the SARS-CoV-2 spike protein (Coutard et al., 2020; Walls et al., 2020). Efficacious SARS-CoV-2 vaccine development is dependent on a robust humoral immune response targeting the spike glycoprotein and, perhaps more specifically, the RBD, assuming similar results as observed for SARS-CoV (Buchholz et al., 2004).

A central feature of the adaptive immune response is presentation of immunogenic peptides for immunosurveillance by



Table 1. Donor Characteristics

| Donor ID | Donor Number | Age | Gender | DRB1 | DRB3,4,5 | USA DRB1 Population Coverage ^a | Asia and Pacific Islands DRB1 Population Coverage ^a |
|----------|--------------|-----|--------|--------------|------------------------|---|--|
| A | 39290 | 33 | male | 07:01, 15:01 | DRB4*01:01, DRB5*01:01 | 23.6% | 16.1% |
| B | 39599 | 40 | female | 07:01, 11:04 | DRB3*02:02, DRB4*01:01 | 14.9% | 8.9% |
| C | 39626 | 29 | female | 15:02, 15:03 | DRB5*01:01, DRB5*01:02 | 2.9% | 8.2% |
| D | 39653 | 35 | male | 07:01, 08:01 | DRB4*01:01 | 14.0% | 8.7% |
| E | 40127 | 31 | female | 04:04, 04:05 | DRB4*01:01 | 4.4% | 6.7% |
| F | 40146 | 41 | female | 07:01, 14:01 | DRB3*02:02, DRB4*01:01 | 14.5% | 10.6% |
| G | 40606 | 52 | male | 04:05, 13:02 | DRB3*03:01, DRB4*01:01 | 5.3% | 9.4% |
| H | 40817 | 21 | male | 13:01, 13:02 | DRB3*02:02, DRB3*03:01 | 9.9% | 6.0% |
| I | 42632 | 57 | female | 11:01 | DRB3*02:02 | 5.8% | 5.1% |
| Total | | | | | | 53.1% | 45.7% |

See Table S1 for extended HLA-II typing.

^a.HLA haplotype frequency data are from the National Marrow Donor Program (<https://bioinformatics.bethematchclinical.org/workarea/downloadasset.aspx?id=9373>).

CD4+ helper T cells. Computational prediction of T cell epitope candidates has been applied to vaccine discovery and removal of unwanted immune responses against protein therapeutic agents (Griswold and Bailey-Kellogg, 2016). Current knowledge of peptide binding motifs is based primarily on data generated using biochemical binding assays (Justesen et al., 2009; Sidney et al., 2013), which are compiled in the Immune Epitope Database (IEDB) (Vita et al., 2019). This information is used to train prediction algorithms, such as the Tepitool resource in the IEDB (Paul et al., 2016) and NetMHCIIpan4 (Reynisson et al., 2020). Recently, this approach was applied to identify known epitopes from multiple coronaviruses and predict likely B and T cell epitopes from several SARS-CoV-2 proteins (Grifoni et al., 2020a), and reactive CD4+ T cells obtained from individuals with COVID-19 were observed when stimulated using a “megapool” comprised of overlapping synthetic 15-mer peptides spanning the entire spike protein sequence (Grifoni et al., 2020b). However, the precise sequences from the SARS-CoV-2 spike glycoprotein for CD4+ T cell activation is lacking, and a clear understanding will benefit vaccine development.

Mass spectrometry-based approaches, often called immunopeptidomics, have been developed to examine the repertoires of peptides presented by human leukocyte antigen (HLA) molecules of major histocompatibility complex (MHC) class I or class II used for immunosurveillance by CD8+ or CD4+ T cells, respectively (Hunt et al., 1992a, 1992b). HLA class II molecules are restricted to antigen-presenting cells (APCs) and play an essential role in development of a humoral adaptive immune response via activation of CD4+ T helper cells. The HLA-II peptide repertoire is a product of extracellular proteins that are proteolytically processed in the lysosomal compartment after internalization. HLA-II immunopeptidomics have been applied to understand the CD4+ T cell epitopes for potential vaccine design from pathogens such as *Mycobacterium tuberculosis* (Bettencourt et al., 2020), vaccinia virus (Strug et al., 2008; Lorente et al., 2019), measles virus (Ovsyannikova et al., 2003), and human herpesvirus 6B (Becerra-Artiles et al., 2019). These approaches have been typically limited in scope, restricted to one or two cell lines, thus sampling only a very limited subset of HLA-DR alleles.

MHC-associated peptide proteomics (MAPPs) is a specific extension of HLA-II immunopeptidomics that incorporates intentional pulsing of dendritic cells (DCs) with an antigen or protein of interest (Röhn et al., 2005). More recently, HLA-II MAPPs has been implemented to investigate and understand the mechanisms of treatment-emergent immunogenicity for biotherapeutic proteins (Walsh et al., 2020; Cassotta et al., 2019; Hamze et al., 2017; Jankowski et al., 2019; Sekiguchi et al., 2018) because this approach allows facile interrogation of the immunogenic potential from multiple HLA class II alleles.

In the present study, we sought to identify the naturally processed and presented immunopeptidome of the SARS-CoV-2 spike glycoprotein from human APCs. DCs from a panel of healthy human subjects representing a large percentage of HLA-DRB1 allele usage within the United States were treated with the recombinant spike glycoprotein extracellular domain (ECD). A subset of donors was also selected to represent common alleles from the Asia-Pacific geographic region. HLA-II-associated peptides were identified by liquid chromatography and nanoelectrospray ionization tandem mass spectrometry after immunoprecipitation. We observed several clusters, or nested sets of peptides, derived from every domain of the SARS-CoV-2 spike glycoprotein. We determined the prevalence of these clusters among multiple donors. Finally, we sought to compare our observed HLA-II epitopes with those from a recent *in silico* prediction (Grifoni et al., 2020a) and determine regions that are conserved with the SARS-CoV spike protein sequence.

RESULTS

DCs from All Donors Display HLA-II SARS-CoV-2 Spike Peptides

The MAPPs method intentionally pulses human DCs from a panel of donors with a protein of interest. The 4 male and 5 female healthy donors used in this study, ranging in age from 21–57 years, were selected to sample approximately 53% and 46% of the HLA-II DRB1 allele frequency from the United States and Asia-Pacific geographic regions, respectively (Table 1). Full HLA typing of the donors is also available (Table S1). The donors’

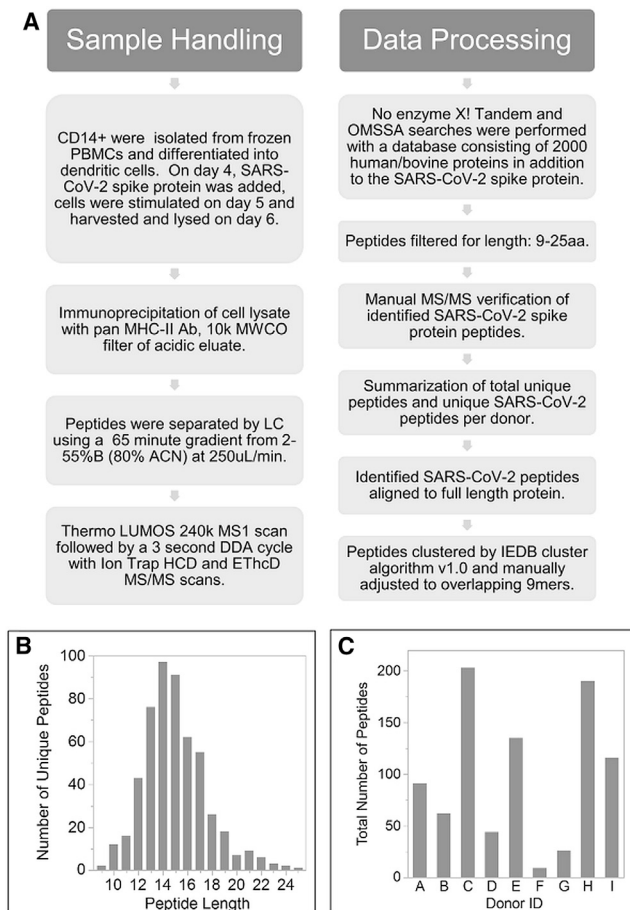


Figure 1. The MAPPs Method and Properties of Observed SARS-CoV-2 Spike Protein-Derived Peptides

(A) Flow diagram summarizing the cell manipulation, immunoprecipitation, and liquid chromatography-tandem mass spectrometry (LC-MS/MS) methods (left) and data processing steps (right).

(B) Distribution of the identified HLA-II peptide lengths from the SARS-CoV-2 spike protein ECD. The size of the peptide is denoted on the x axis, and the number of unique peptides is denoted on the y axis.

(C) Bar graph of the total number of SARS-CoV-2 spike protein ECD peptides identified per donor. The donor ID is listed on the x axis, and the total number of spike protein peptides from that particular donor is denoted on the y axis.

See [Tables 1](#) and [S1](#) for donor information and [Table S2](#) for all identified SARS-CoV-2 spike protein ECD peptides.

PBMCs were collected and stored frozen before the outbreak of the SARS-CoV-2 pandemic and are expected to be unexposed to potential infection. The sequence used for this work was derived from the SARS-CoV-2 Wuhan-Hu-1 strain spike glycoprotein ECD spanning residues 1–1,213 taken from GenBank: YP_009724390. An R685A mutation was used to prevent cleavage during recombinant protein production, with affinity tags added for purification. The recombinant protein was produced from mammalian cells and is expected to have N-linked glycosylation modifications.

The method used to profile the HLA-II peptides is outlined in [Figure 1A](#). Monocyte-derived DCs were generated in culture with a cytokine cocktail. The immature DCs were treated

with the SARS-CoV-2 spike glycoprotein ECD, and after 24 h, lipopolysaccharide (LPS) was added to mature the DCs. The treated cells were lysed, and the HLA class II complex was isolated by immunoprecipitation with a pan-HLA class II antibody. The bound HLA-II peptides were eluted after acidification, filtered to remove high-molecular-weight co-precipitants, and analyzed by capillary HPLC on an orbitrap mass spectrometer. Peptide ions were fragmented using multiple fragmentation techniques. Peptides were identified using multiple proteomic search engines with a forward and reverse database search. False discovery rate estimates (q values) were estimated using a null distribution from the reverse database search. The peptides identified from the spike glycoprotein had q values below 0.07, and all fragmentation spectra matching the spike protein were reviewed manually (see [STAR Methods](#) for details).

A total of 876 HLA-II peptides from the SARS-CoV-2 spike glycoprotein were identified from the donor panel ([Table S2](#)). Several peptides with identical sequences were identified from multiple donors. Removal of these duplicate peptides resulted in 526 unique sequences. To minimize false positive and negative peptide identification, the database used for peptide identification was intentionally constructed to contain the spike protein along with approximately 2,000 background bovine and human proteins observed previously from multiple samples analyzed with this assay system. We also analyzed the data using multiple search engines against a database containing the entire human proteome (downloaded on April 5, 2019) that also included the SARS-CoV-2 spike glycoprotein and did not find any human protein identification for the 526 unique SARS-CoV-2 spike peptides. The primary mass spectrometry data are publicly available for individual analysis (<https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000085456>).

The unique peptide sequences had a distribution of lengths consistent with HLA-II peptides, with a mean length of 15 residues ([Figure 1B](#); [Kampstra et al., 2019](#)). There were 169 unique peptides that had modified residues. The modifications observed were consistent with HLA-II peptide processing. The spike protein is heavily glycosylated with 22 putative N-glycosylation sites, presumably to help evade immune detection, and the majority of the regions from the spike ECD not observed were centered around these sites. Current limitations of our method prevent detection of glycosylated peptides if, in fact, they were processed and loaded onto HLA class II molecules. Interestingly, HLA-II peptides from 3 putative N-linked glycosylation sites were observed, which indicates that these regions were not modified. We did identify a deamidation modification with the asparagine residue at residue 1,098 with our search engines ([Table S2](#)). This modification is consistent with conversion of asparagine to aspartic acid after removal of N-linked glycosylation and is indirect evidence of glycosylation of this residue in the intact protein ([Khoshnoodi et al., 2007](#)). Every donor examined produced HLA-II peptides derived from the SARS-CoV-2 spike glycoprotein ([Figure 1C](#)). The number of spike peptides observed per donor ranged from 9–203 with a median value of 91 peptides. One donor (40146) did not efficiently produce mature DCs and resulted in a low overall number of peptides relative to all other donors

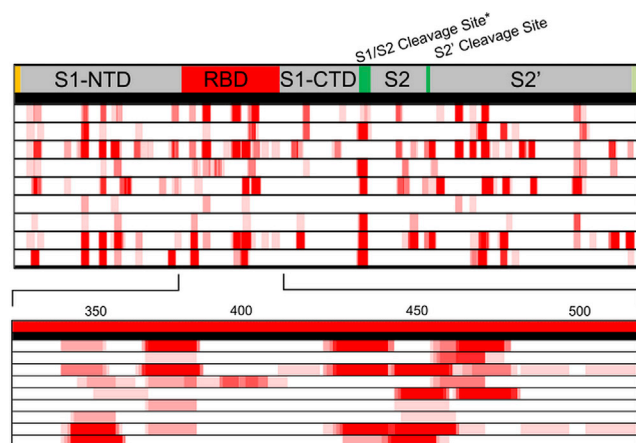


Figure 2. HLA-II Peptides Derived from the SARS-CoV-2 Spike Protein ECD

The various subunits of the spike protein are denoted at the top of the schematic. Aligned HLA-II-presented peptides are displayed as a heatmap with each of the nine donors in an individual row. The shades of red correspond to the number of overlapping peptides encompassing a given amino acid position. The lightest shade represents a single peptide, and the darkest red signifies when at least five peptides overlap an amino acid position. The unlabeled yellow region at the N terminus corresponds to the signal peptide, and the light green region at the C terminus corresponds to the affinity tag used for purification. The S1/S2 cleavage site contains an R685A mutation (indicated by an asterisk). The RBD portion of the heatmap is expanded and displayed in the bottom panel along with numerical markers to indicate location within the SARS-CoV-2 spike protein ECD. Data from each donor were collected from a single biological and technical replicate. See [Table S2](#) for all identified SARS-CoV-2 spike protein ECD peptides. S1, subunit 1; S2, subunit 2; NTD, N-terminal domain; RBD, receptor binding domain; CTD, C-terminal domain.

analyzed. Repeated analysis with this donor also yielded poor results (data not shown).

HLA-II Peptides Are Distributed across the Entire SARS-CoV-2 Spike Protein ECD and Have Consensus HLA-II Clusters

HLA-II peptides derived from the SARS-CoV-2 spike glycoprotein were aligned to the ECD sequence and a peptide density map was generated to help visualize the breadth and depth of sequence coverage. A schematic listing of the various subunits of the spike glycoprotein ECD and the results obtained with each donor are shown in individual rows ([Figure 2](#)). The shades of red correspond to the number of overlapping peptides encompassing a given amino acid position. From this presentation of the data, it is evident that HLA-II peptides spanning all segments of the spike glycoprotein ECD were obtained from all sampled donors. The heatmap also reveals that HLA-II peptides from several regions of the spike glycoprotein were observed from multiple donors, likely reflecting the more promiscuous HLA class II binding epitopes. An expanded view of the RBD encompassing residues 319–541 is shown in the bottom panel of [Figure 2](#) and clearly demonstrates that promiscuous epitopes are also contained in this region of the molecule, critical for interaction with the ACE2 cell receptor.

Groups of nested peptides sharing a common core but with ragged N and C termini are generated from the multiple proteases and different temporal patterns of processing that occur in the

lysosomal compartment ([Lippolis et al., 2002](#)). To organize the observed HLA-II peptides from the spike protein into discrete segments for subsequent analysis, we used the IEDB Epitope Cluster Analysis Tool 1.0 with some manual adjustments, as noted in [STAR Methods](#), to group these into 73 distinct clusters ([Table S3](#)). Clusters were characterized from full-span and minimal-overlap perspectives. The full-cluster sequence represents the first start position of the peptides in the cluster to the last position of the peptides in the cluster. The minimum-cluster sequence is the smallest common sequence among the peptides in the cluster. Clusters with minimum-cluster sequences of less than 9 residues likely contain 2 overlapping binding cores that, because of their proximity, were unable to be separated easily.

As expected, we observed a distribution of the clusters among the donors ([Figure 3](#)). Most of the clusters were observed from 4 or fewer donors and were designated as restricted, whereas 11 clusters were observed from 5–7 donors and were designated as consensus ([Figure 3A](#); [Table 2](#)). This arbitrary definition of a consensus cluster was set to reflect sequences observed in at least 50% of the donors sampled in the study. Consensus clusters represent SARS-CoV-2 spike glycoprotein sequences with the most promiscuous but not necessarily highest affinity, binding to the broadest range of HLA class II alleles. No specific cluster was observed in all donors sampled in this study. The median number of clusters per donor was 18, and all donors displayed at least one of the consensus clusters ([Figure 3B](#)). We next examined the distribution in the number of peptides in restricted and consensus clusters ([Figure 3C](#)). The vast majority of clusters contained less than 10 peptides, with most but not all of the consensus clusters having 11 or greater members. Interestingly, consensus cluster 1, observed in 5 donors, was comprised of a single peptide sequence in the S1 subunit ([Table 2](#)). Finally, we examined the location of the clusters within discrete segments of the SARS-CoV-2 spike glycoprotein ([Figure 3D](#)). The clusters were distributed throughout the protein. All regions contained at least one HLA-II peptide cluster, with consensus clusters occurring in several different regions. Of note, the RBD region responsible for binding ACE2 and a target for vaccination strategies contained a total of 16 clusters in which 3 were consensus ([Table 3](#)).

Observed SARS-CoV-2 Spike Protein HLA-II Peptides Have Limited Overlap with Predicted CD4+ T Cell Epitopes

CD4+ T cell epitopes from the spike glycoprotein derived from an algorithm designed to predict the dominant HLA class II peptides were published recently ([Grifoni et al., 2020a](#)). We sought to compare those predictions with the results from our study. We used the minimum observed cluster sequence for our comparison and considered an overlap of at least 9 residues within the predicted 15 residue peptide sequence as a match, with one exception, as denoted below. We chose this minimum overlap length based on the number of residues contained in the HLA class II peptide binding cleft. We chose to use the minimum-cluster sequence in an attempt to reduce matches on the extreme peripheries of the observed peptides that likely do not reflect the likely binding core contained in the cluster.

A perfect congruence between prediction and observed results from this donor set would result in 19 matches as one of

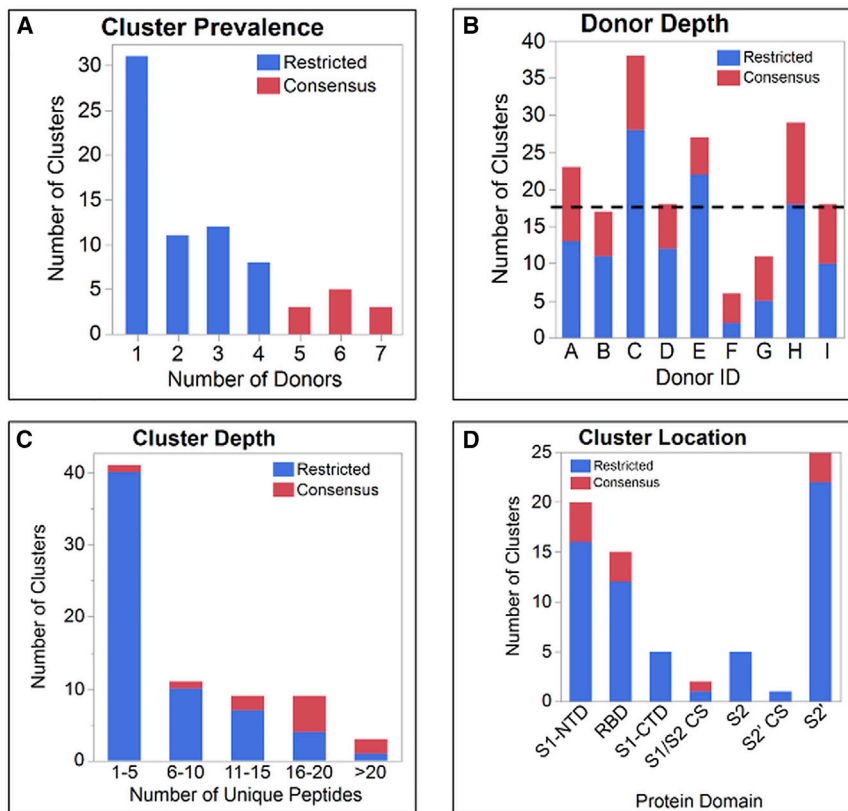


Figure 3. Properties of Observed HLA-II Clusters from the SARS-CoV-2 Spike Protein ECD

(A) Prevalence of peptide clusters across donors. The number of donors from whom each of the 73 peptide clusters were observed are shown. The x axis denotes the number of donors in whom a cluster was observed, and the y axis denotes the number of clusters observed from that particular number of donors. For example, 31 clusters were observed from any given single donor, another 10 clusters were observed from 2 of the 9 donors in the panel, and so forth. Peptide clusters that were identified in at least 5 donors were deemed “consensus” and are colored red, whereas clusters seen in 4 or less donors are characterized as “restricted” and are colored blue.

(B) Number and type of cluster by donor. The total numbers of clusters observed for each donor are shown. The y axis denotes the total number of clusters from the particular donor listed on the x axis. Clusters designated as restricted or consensus are shown in blue and red, respectively. The overall median of 18 clusters is indicated by a dashed line.

(C) Cluster depth. The distribution in the different number of peptides contained in a cluster is shown. The number of peptides in any given cluster is denoted on the x axis, and the number of clusters containing the peptides in the designated bins is denoted on the y axis. Clusters designated as restricted or consensus are shown in blue and red, respectively.

(D) The distribution of the clusters across the spike

protein domains. The particular domain is designated on the x axis, and the number of clusters contained in the particular domain is denoted on the y axis. Clusters designated as restricted or consensus are shown in blue and red, respectively. See [Table S3](#) for further cluster details.

the predicted peptides resides in the transmembrane portion that was absent from the protein used in this study. In contrast, we observed HLA-II peptide clusters from our donor set that matched a total of 9 predicted epitopes ([Figure 4A](#); [Table S4](#)). Cluster 27 was deemed to match a predicted epitope located from residues 451–465 (YLRLFRKSNLKPFE) in the RBD domain even though only a 5-residue overlap was observed using the criteria defined above. This particular cluster was composed of 37 unique sequences, the most of all clusters, and likely contains two closely overlapping allele binding sites. Because 25 of the observed peptides from this cluster matched the first 9 residues of the predicted peptide, we included this region as overlapping with the prediction (see [Table S2](#) for cluster details). We also noted two instances in which multiple clusters were deemed to sufficiently overlap with a predicted peptide, which skews the Venn diagram in [Figure 4A](#) to show 71 instead of the 73 total observed clusters.

Comparison of the observed consensus clusters with the predicted peptides reveals that only 2 of the 11 observed promiscuous HLA-II binding regions were predicted ([Table 2](#)). This finding is intriguing, given the prevalence of display of these particular sequences, and may reflect limitations in the alleles used to generate the predictions and/or the prediction algorithm. The RBD of the SARS-CoV-2 spike glycoprotein contained 16 observed HLA-II clusters and 4 predicted peptides. We

observed 3 of the 4 predicted epitopes ([Tables 3](#) and [S3](#)). The only predicted peptide that was not observed contained a putative N-linked glycosylation site. As denoted above, our MAPPs method is unlikely to identify a peptide with this modification. In fact, 6 of the 20 predicted peptides contain a putative N-linked glycosylation site ([Table S4](#)), possibly reflecting the fact that current predictive algorithms do not consider post-translational modifications. In addition, the glycosylated asparagine residue in many of the predicted peptides is located centrally rather than on the periphery, and large complex glycan structures have been shown to interfere with HLA or T cell receptor binding ([Speir et al., 1999](#); [Kario et al., 2008](#); [Malaker et al., 2017](#)). Also striking is the overall number of observed clusters (62), many obtained from multiple donors, which were not predicted. It is worth noting that the algorithm used to predict SARS-COV-2 spike protein epitopes is largely trained on HLA binding affinity data that do not reflect the authentic processing captured by the MAPPs assay. Further expansion of MAPPs-derived data into these training sets will likely be beneficial.

Sequence Analysis of Observed HLA-II Peptides Show Limited Overlap with SARS-CoV and No Match with Other CoV Spike Proteins or Any Human Sequences

The list of unique peptides for each HLA class II cluster was compared with the spike glycoprotein sequence from SARS-

Table 2. SARS-CoV-2 Spike Protein ECD Consensus Clusters

| Cluster ID | Number of Donors | Number of Unique Peptides | Protein Domain | Start | End | Full Cluster Sequence | Predicted ^a | SARS-CoV |
|------------|------------------|---------------------------|----------------|-------|-------|--|------------------------|----------|
| 1 | 5 | 1 | S1-NTD | 24 | 40 | LPPAYTNSFTRGVVYPD | | |
| 6.2 | 7 | 16 | S1-NTD | 130 | 148 | VCEFQFCNDPFLGVYYHKN | | |
| 7 | 7 | 17 | S1-NTD | 166 | 182 | CTFEYVVSQPFLLMDLEGK | | |
| 8.2 | 6 | 16 | S1-NTD | 195 | 218 | KNIDGYFKIYSKHTPINLVRDLPO | | |
| 18.1 | 5 | 12 | RBD | 344 | 363 | ATRFASVYAWNKRKRISNCVA | | |
| 21 | 6 | 19 | RBD | 367 | 387 | VLYNSASFSTFKCYGVSPSTKL | | |
| 27 | 6 | 37 | RBD | 443 | 466 | SKVGGNYNYLRYLFRKSNLKPFFER | X | |
| 37 | 7 | 24 | S1/S2 CS | 672 | 696 | ASYQTQTNSPRRAA ^b SVASQSIAYT | | |
| 49 | 6 | 6 | S2' | 894 | 910 | LQIPFAMQMAYRFNGIG | X | X |
| 52 | 5 | 11 | S2' | 917 | 935 | YENQKLIANQFNSAIGKIQ | | |
| 64 | 6 | 20 | S2' | 1,096 | 1,116 | VSNQTHWFVTQRNFYEPQIIT | | |

NTD, N-terminal domain; RBD, receptor binding domain; CS, cleavage site.

^aGrifoni et al. (2020a).

^bR685A mutation used for production of recombinant protein.

CoV protein (UniProt: P59594). The non-identical residues between the two sequences are shown in red under the spike glycoprotein schematic in Figure 4B. The S2 and S2' regions had larger areas of identity between the sequences. The 526 unique HLA-II peptides identified in this study were analyzed for sequence identity with the SARS-CoV spike glycoprotein. No identical matches were identified in the S1 subunit. However, 14 clusters match the SARS-CoV sequence in the S2 subunit. Interestingly, one of the matches, cluster 49, was a consensus cluster (denoted in blue Figure 4B). These clusters represent sequence regions from both viruses that are potentially presented by HLA class II molecules for T cell surveillance. The spike proteins from the other coronaviruses known to infect humans (NL63, 229E, HKU1, OC43, and MERS) were also evaluated, and no matches were found.

To look for potential cross-reactivity to human proteins, each unique identified peptide from the SARS-CoV-2 spike glycoprotein was searched against the UniProt human database for an exact match to any human protein. None of the observed peptides had a sequence match (data not shown). Finding no matches, we expanded our search to include up to 2 mismatched amino acids without insertions or deletions. A single peptide could be associated with 13 human proteins when 2 residues of non-identity are allowed. These results indicate that the risk from direct sequence cross-reactivity is minimal and that any portion of the SARS-CoV-2 spike glycoprotein associated with HLA class II molecules is unlikely to be subject to previous tolerization in a vaccinated subject or infected individual. However, we acknowledge that our analysis does not consider cross-reactivity when strictly limited to putative T cell contact residues, given the difficulty to reliably predict such registers.

DISCUSSION

CD4+ T cell participation is vital for a robust humoral response to viral infection or active immunization. A clear delineation of the epitopes presented by APCs for T cell immunosurveillance greatly enhances our understanding of this process. Generally,

T cell lines or PBMCs from recovered patients using peptides derived simply by spanning the entire protein(s) of interest or from HLA-II prediction algorithms have been utilized (Meunier et al., 2019; Grifoni et al., 2020b; Schulze zur Wiesch et al., 2005). The ability to automate and miniaturize the MAPPs assay enables facile identification of thousands of naturally processed and displayed HLA-II peptides from human DCs. Using this approach, we were able to determine the precise regions and sequences of peptides from the SARS-CoV-2 spike glycoprotein ECD, derived from a panel of healthy subjects, presented for immune surveillance by T cells. The 9 subjects used in this study enabled sampling of approximately 53% and 46% of the HLA-DRB1 allele frequency from the United States and Asia-Pacific geographic regions, respectively (Table 1). To our knowledge, this work represents the most precise and comprehensive immunopeptidomic investigation with SARS-CoV-2 spike glycoprotein performed to date and allows detailed analysis of features that may aid vaccine development.

We observed a total of 526 unique peptide sequences contained within 73 clusters distributed across each segment of the SARS-CoV-2 spike glycoprotein ECD presented by human DCs (Figure 2; Table S2). Two of the clusters were in regions that deviated from the reference sequence. One region was the S1/S2 cleavage site, in which the novel furin site was eliminated with mutation to enable production of full-length recombinant protein. We speculate that this region of the spike protein containing the native residue could also be presented from molecules that are not cleaved during virion particle assembly. The other area of deviation was the C-terminal affinity tags used for purification. Of particular interest are peptides from the spike glycoprotein that are presented by multiple donors because these would be sequences likely to elicit a T cell response from the greatest number of patients or vaccinated subjects. We observed 11 consensus clusters, defined as being present in 5 or more of the 9 donors analyzed in this study, including 3 in the RBD, which is essential for binding to ACE2 on host cells (Table 3). A majority of the consensus clusters contained 11 or more nested peptides. In the absence of a dedicated assay to quantify

Table 3. SARS-CoV-2 Spike Protein ECD RBD HLA-II Clusters

| Cluster ID | Number of Donors | Number of Unique Peptides | Start | End | Full Cluster Sequence | Predicted ^a |
|------------|------------------|---------------------------|-------|-----|-------------------------|------------------------|
| 18.1 | 5 | 12 | 344 | 363 | ATRFASVYAWNRKRISNCVA | |
| 18.2 | 4 | 9 | 348 | 363 | ASVYAWNRKRISNCVA | |
| 19 | 1 | 1 | 352 | 366 | AWNRKRISNCVADYS | |
| 20 | 1 | 1 | 354 | 370 | NRKRISNCVADYSVLYN | |
| 21 | 6 | 19 | 367 | 387 | VLYNSASFSTFKCYGVSPKTL | |
| 22 | 1 | 1 | 382 | 398 | VSPTKLNLCFTNVYAD | |
| 23 | 1 | 3 | 393 | 407 | TNVYADSFVIRGDEV | |
| 24 | 1 | 1 | 401 | 413 | VIRGDEVQRQIAPG | |
| 25 | 1 | 1 | 411 | 423 | APGQTGKIADYNY | |
| 26 | 4 | 17 | 425 | 446 | LPDDFTGCVIAWNSNNLDSKVG | |
| 27 | 6 | 37 | 443 | 466 | SKVGGNYLYRLFRKSNLKPFFER | X |
| 28 | 4 | 4 | 458 | 474 | KSNLKPFFERDISTEIYQ | X |
| 29 | 4 | 17 | 465 | 485 | ERDISTEIYQAGSTPCNGVEG | |
| 30 | 2 | 1 | 486 | 500 | FNCYFPLQSYGFQPT | |
| 31 | 2 | 1 | 506 | 522 | QPYRVVLSFELLHAPA | X |
| 32 | 1 | 4 | 544 | 558 | NGLTGTGVLTESNKK | |

^aGrifoni et al. (2020a).

the presented HLA-II peptides, we use this metric as a surrogate for peptide abundance but recognize that even a single specific peptide sequence can be presented in sufficient numbers to elicit a T cell response.

Recent reports leveraging bioinformatics to predict (Grifoni et al., 2020a) or a single-pot peptide pool composed of more than 150 overlapping peptides spanning the entire open reading frame (Grifoni et al., 2020b; Braun et al., 2020) have been published, attempting to elucidate SARS-CoV-2 T cell epitopes. Unfortunately, the latter approach does not allow any insights into the precise sequences capable of eliciting a response. Comparisons of the clusters we observed being presented by APCs, reflecting the natural HLA-II loading processes, with the predicted epitopes is illuminating. One of the predicted epitopes resides in the transmembrane domain that was absent from the protein used for our analysis and omitted from further discussion. Of the remaining predictions, roughly 50% (9 of 19) were observed from our panel of donors selected to represent a sizeable percentage of HLA-DR allele usage from multiple geographic regions (Figure 4; Table S3). Correspondingly, the vast majority of the observed 73 HLA-II clusters were not predicted. Of particular interest are the consensus clusters that were observed in 5 or more of the donors and would be expected to represent sequences with the most promiscuous HLA class II binding. Only 2 of these 11 consensus clusters were predicted, with only 1 of 3 consensus clusters contained in the RBD predicted. Of note, and expanded in more detail below, consensus cluster 49 in the S2 subunit has 100% sequence identity to SARS-CoV, was predicted, and has been experimentally shown to be a T cell epitope (Yang et al., 2009).

The “7-allele method” HLA class II reference set used for generating the predicted epitopes is restricted to select DRB1/3/4/5 alleles (<http://tools.iedb.org/mhcii/>; Paul et al., 2015). The use of the pan-HLA class II antibody used in our study, which would enrich HLA-DQ- and HLA-DP-bound peptides, could

explain why some but certainly not all of the observed clusters were not predicted. However, given ample evidence that shows the effect of HLA-DQ- and HLA-DP-bound peptides on T cell activation for a variety of viral antigens (Koelle et al., 1997; Mellins et al., 1987; Koelle et al., 1994; Lorente et al., 2019, 2020), we thought that it was important to identify as many of those restricted peptides as possible. Nevertheless, this disconnect between promiscuously observed HLA class II clusters and predicted T cell epitopes accentuates what has been highlighted before: that prediction of T cell epitopes is an imperfect process that may not reflect what HLA class II molecules bind preferentially (Paul et al., 2015; Wantuch et al., 2020) and, therefore, that this is not the most effective approach for identifying CD4+ T cell epitopes.

At present, we have not confirmed whether all or some subset of the SARS-CoV-2 spike protein HLA-II peptides presented by DCs are actual CD4+ T cell epitopes and cause cell proliferation. Multiple examples indicate that the sequences identified will be *bona fide* epitopes because peptides identified by mass spectrometry methods similar to the work presented here have been shown to elicit recall T cell responses from vaccinated individuals and individuals with active viral infection (Becerra-Artiles et al., 2019; Wantuch et al., 2020; Strug et al., 2008; Ovsyanikova et al., 2003). MAPPs has been an instrumental method to preclinically assess the immunogenic potential of protein therapeutic agents (for a review, see Quarmby et al., 2018). Multiple reports analyzing different therapeutic antibodies approved for clinical use have shown that many but not all HLA-II clusters identified from these molecules can elicit T cell responses from drug-naïve donors as well as individuals who developed treatment-emergent anti-drug antibody responses (Walsh et al., 2020; Cassotta et al., 2019; Hamze et al., 2017). Also, unlike the immunogenicity assessment of a majority of protein biotherapeutic agents, which are engineered to a great extent to be recognized as a self-protein and, therefore, generally present

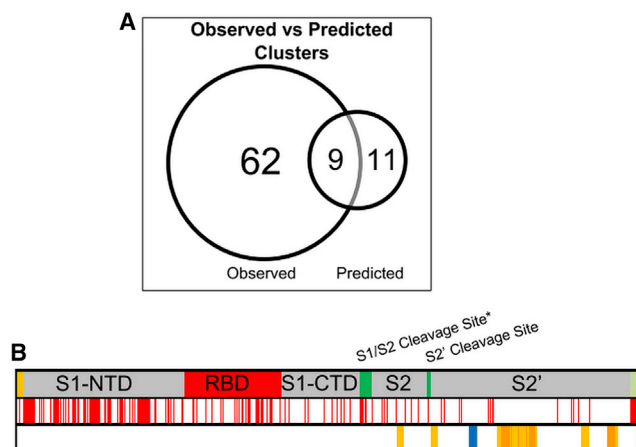


Figure 4. Observed SARS-CoV-2 Spike Protein HLA-II Peptide Clusters versus Predictions and Similarity to SARS-CoV

(A) Prediction comparison. The observed SARS-CoV-2 spike protein clusters were compared with the predicted HLA-II peptides. A Venn diagram shows the overlap between the observed minimum cluster sequence and the predicted peptides. To view the details of predicted versus observed clusters, see [Table S4](#).

(B) Cluster conservation. The various segments of the SARS-CoV-2 spike protein are denoted on the top row, and sequence mismatches to the SARS-CoV spike protein are delineated by a red bar in the second row. The third row is a heatmap of the clusters seen from the SARS-CoV-2 spike protein that contain peptides with an exact sequence match to the SARS-CoV spike protein (darker orange represents overlapping clusters). The blue cluster designates a consensus cluster.

a very limited number of non-germline residues for scrutiny, each SARS-CoV-2 spike protein cluster is fundamentally distinct from any other sequence in the human genome and extremely unlikely to be subject to any tolerization. The list of unique SARS-CoV-2 spike glycoprotein peptides identified by MAPPs searched against the human database (UniProt version 2020_2) did not identify any significant matches. Nevertheless, the activation potential of the SARS-CoV-2 spike protein clusters with T cells from healthy donors and, ideally, convalescent individuals should be evaluated.

The DCs used were derived from healthy donors who, because of the timing of sample collection, had not been exposed to potential SARS-CoV-2 infection. Therefore, the displayed HLA-II clusters reported here could conceivably deviate from the repertoire obtained from infected individuals. This potential discrepancy would require that internalization of the virus into APCs would fundamentally alter the proteolysis and/or HLA-II molecule loading mechanism. Although examples of HLA-II display interference are known for other viruses that can infect and replicate in immune cells ([Becerra-Artiles et al., 2019](#)), we are not aware of that attribute with SARS-CoV-2 at this time and do not consider this to be a fundamental concern. This could be addressed by repeating this study using material obtained from convalescent individuals or infection of DCs from healthy donors with live virus. Also, the DCs used in this study were derived from monocytes and could potentially have a different processing mechanism from plasmacytoid or follicular DC lineages. A detailed comparison in the MAPPs peptides obtained

from different DC types is lacking, given the difficulties in obtaining adequate numbers required for investigation. Nevertheless, multiple studies have shown monocyte-derived DCs to be as efficient as antigen-specific B cells in presenting peptides for T cell surveillance ([Cella et al., 1997](#); [Sallusto and Lanzavecchia, 1994](#)).

Insights into the immunogenic potential of SARS-CoV-2 spike glycoprotein can be made from the results obtained in this study. First, the depth and breadth of the HLA-II peptides derived from this critical structural component for viral infectivity indicate that mutational drift as the pandemic continues to spread around the world is not expected to dramatically alter the ability of an infected individual to mount a new B cell response to replace antibodies that would be affected deleteriously by such escape mutations. If a mutation could result in an inability of a particular cluster to bind most HLA class II alleles, then the sheer number of clusters distributed across the spike protein, especially the consensus clusters, make the mutation unlikely to enable the virus to escape CD4+ T cell activation across a wide portion of the population. Second, SARS-CoV spike glycoprotein cross-reactive HLA-II-restricted epitopes seem to be limited to the S2 domain because all 14 of the SARS-CoV-2 clusters identified in this study with complete sequence identity to SARS-CoV were derived from this region of the molecule ([Figure 4B](#); [Table S3](#)). This result is not surprising, given the 90% sequence identity between the two viruses in S2 but only 60% identity in the S1 domain. No significant homology was noted for any of the other human coronavirus spike protein sequences. Interestingly, consensus cluster 49 (observed from 6 donors) was one of the 5 clusters from the S2 subunit with complete identity to SARS-CoV and has been identified previously as a T cell epitope from healthy donors using a tetramer-guided epitope mapping approach ([Yang et al., 2009](#)). This epitope has significant, but not complete, overlap with the HLA-II derived cluster and indicates, as denoted above, that the approach of using overlapping peptides may reflect imperfect identifications from those that arise because of natural APC processing.

The limitations of this study include that the identified peptides are restricted to those with suitable biophysical characteristics for ionization and compatibility with reverse-phase liquid chromatography. The total number and wide breadth of coverage spanning the entire ECD of the spike protein indicate that any false negative results obtained with the method outlined in this study are likely a small minority. Another limit is focusing the analysis on the spike glycoprotein to the exclusion of the other structural components of SARS-CoV-2: the membrane and nucleocapsid proteins. Undoubtedly many regions from those proteins will also be presented for T cell surveillance; however, the focus of most immunization strategies seems to be to target the spike glycoprotein. Notwithstanding, the “adjuvant-like” potential of some of these presumed clusters to augment the humoral response to the spike glycoprotein cannot be accounted for with the current results and should be targets of future efforts. Additional effort to experimentally identify the actual HLA-II peptides presented from the spike glycoprotein, at a minimum, from all other coronaviruses would be of interest.

The ability to direct the humoral immune response to discrete segments of the SARS-CoV-2 spike glycoprotein that confer

viral neutralization may potentially enable higher protective titers to be achieved with vaccination and limit antibody-dependent enhancement of infection, as reported with other coronaviruses (Tseng et al., 2012; Wan et al., 2020). A preliminary report in which the 197-amino-acid RBD of the SARS-CoV-2 spike glycoprotein was used for immunization in rodents suggests that a robust neutralizing response can be obtained without antibody-dependent enhancement (Quinlan et al., 2020). The outlined approach and the results reported in this study can also be applied to developing novel subunit or nucleic acid-based vaccines and/or monitoring the response to such vaccines. It also enables the ability to supply the immune system with synthetic peptide(s) that mirror natural APC presentation observed from a broad spectrum of HLA class II alleles from different geographic regions to maximize T cell responses.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
 - Dendritic Cell Culture
 - HLA-II Peptide Isolation
 - LC/MS analysis of HLA-II Eluted Peptides
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108454>.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Richard E. Higgs, Andrea Ferrante, and Laurent Malherbe for critical reviews and helpful suggestions during preparation of the manuscript. R.W.S. dedicates this work to the memory of Karen J. Haugh. This work was supported by Eli Lilly and Company.

AUTHOR CONTRIBUTIONS

Supervision, Software, Visualization, Formal Analysis, Data Curation, and Writing – Original Draft, M.D.K.; Investigation, Software, Formal Analysis, Visualization, and Writing – Original Draft, M.B.L.; Resources, Investigation, Visualization, and Writing – Original Draft, L.J.S.; Resources, Supervision, and Writing – Original Draft, R.W.S.; Writing – Review & Editing, R.J.K. and C.L.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 24, 2020
Revised: October 16, 2020
Accepted: November 6, 2020
Published: December 1, 2020

REFERENCES

- Becerra-Artiles, A., Cruz, J., Leszyk, J.D., Sidney, J., Sette, A., Shaffer, S.A., and Stern, L.J. (2019). Naturally processed HLA-DR3-restricted HHV-6B peptides are recognized broadly with polyfunctional and cytotoxic CD4 T-cell responses. *Eur. J. Immunol.* *49*, 1167–1185.
- Bettencourt, P., Müller, J., Nicastrì, A., Cantillon, D., Madhavan, M., Charles, P.D., Fotso, C.B., Wittenberg, R., Bull, N., Pinpathomrat, N., et al. (2020). Identification of antigens presented by MHC for vaccines against tuberculosis. *NPJ Vaccines* *5*, 2.
- Braun, J., Loyal, L., Frentsch, M., Wendisch, D., Georg, P., Kurth, F., Hippenstiel, S., Dingeldey, M., Kruse, B., Fauchere, F., et al. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* *587*, 270–274.
- Buchholz, U.J., Bukreyev, A., Yang, L., Lamirande, E.W., Murphy, B.R., Subbarao, K., and Collins, P.L. (2004). Contributions of the structural proteins of severe acute respiratory syndrome coronavirus to protective immunity. *Proc. Natl. Acad. Sci. USA* *101*, 9804–9809.
- Cassotta, A., Mikol, V., Bertrand, T., Pouzieux, S., Le Parc, J., Ferrari, P., Dumas, J., Auer, M., Deisenhammer, F., Gastaldi, M., et al. (2019). A single T cell epitope drives the neutralizing anti-drug antibody response to natalizumab in multiple sclerosis patients. *Nat. Med.* *25*, 1402–1407.
- Cavanagh, D. (1995). The coronavirus surface glycoprotein. In *The coronaviridae*, S.G. Siddell, ed. (Boston, MA: Springer), pp. 73–113.
- Cella, M., Engering, A., Pinet, V., Pieters, J., and Lanzavecchia, A. (1997). Inflammatory stimuli induce accumulation of MHC class II complexes on dendritic cells. *Nature* *388*, 782–787.
- Coutard, B., Valle, C., de Lamballerie, X., Canard, B., Seidah, N.G., and Decroly, E. (2020). The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Res.* *176*, 104742.
- Frank, A., and Pevzner, P. (2005). PepNovo: de novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* *77*, 964–973.
- Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R.H., Peters, B., and Sette, A. (2020a). A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host Microbe* *27*, 671–680.e2.
- Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jardi, R.S., et al. (2020b). Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* *181*, 1489–1501.e15.
- Griswold, K.E., and Bailey-Kellogg, C. (2016). Design and engineering of deimmunized biotherapeutics. *Curr. Opin. Struct. Biol.* *39*, 79–88.
- Hamze, M., Meunier, S., Karle, A., Gdoura, A., Goudet, A., Szely, N., Pallardy, M., Carbonnel, F., Spindeldreher, S., Mariette, X., et al. (2017). Characterization of CD4 T Cell Epitopes of Infliximab and Rituximab Identified from Healthy Donors. *Front. Immunol.* *8*, 500.
- Higgs, R.E., Knierman, M.D., Gelfanova, V., Butler, J.P., and Hale, J.E. (2008). Label-free LC-MS method for the identification of biomarkers. *Methods Mol. Biol.* *428*, 209–230.
- Hunt, D.F., Henderson, R.A., Shabanowitz, J., Sakaguchi, K., Michel, H., Sevilier, N., Cox, A.L., Appella, E., and Engelhard, V.H. (1992a). Characterization of peptides bound to the class I MHC molecule HLA-A2.1 by mass spectrometry. *Science* *255*, 1261–1263.
- Hunt, D.F., Michel, H., Dickinson, T.A., Shabanowitz, J., Cox, A.L., Sakaguchi, K., Appella, E., Grey, H.M., and Sette, A. (1992b). Peptides presented to the immune system by the murine class II major histocompatibility complex molecule I-Ad. *Science* *256*, 1817–1820.
- Jankowski, W., Park, Y., McGill, J., Maraskovsky, E., Hofmann, M., Diego, V.P., Luu, B.W., Howard, T.E., Kellerman, R., Key, N.S., and Sauna, Z.E. (2019). Peptides identified on monocyte-derived dendritic cells: a marker for clinical immunogenicity to FVIII products. *Blood Adv.* *3*, 1429–1440.

- Justesen, S., Hamdahl, M., Lamberth, K., Nielsen, L.L., and Buus, S. (2009). Functional recombinant MHC class II molecules and high-throughput peptide-binding assays. *Immunome Res.* *5*, 2.
- Käll, L., Canterbury, J.D., Weston, J., Noble, W.S., and MacCoss, M.J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* *4*, 923–925.
- Kampstra, A.S.B., van Heemst, J., Janssen, G.M., de Ru, A.H., van Lummel, M., van Veelen, P.A., and Toes, R.E.M. (2019). Ligandomes obtained from different HLA-class II-molecules are homologous for N- and C-terminal residues outside the peptide-binding cleft. *Immunogenetics* *71*, 519–530.
- Kario, E., Tirosh, B., Ploegh, H.L., and Navon, A. (2008). N-linked glycosylation does not impair proteasomal degradation but affects class I major histocompatibility complex presentation. *J. Biol. Chem.* *283*, 244–254.
- Khoshnoodi, J., Hill, S., Tryggvason, K., Hudson, B., and Friedman, D.B. (2007). Identification of N-linked glycosylation sites in human nephrin using mass spectrometry. *J. Mass Spectrom.* *42*, 370–379.
- Koelle, D.M., Corey, L., Burke, R.L., Eisenberg, R.J., Cohen, G.H., Pichyangkura, R., and Triezenberg, S.J. (1994). Antigenic specificities of human CD4+ T-cell clones recovered from recurrent genital herpes simplex virus type 2 lesions. *J. Virol.* *68*, 2803–2810.
- Koelle, D.M., Johnson, M.L., Ekstrom, A.N., Byers, P., and Kwok, W.W. (1997). Preferential presentation of herpes simplex virus T-cell antigen by HLA DQA1*0501/DQB1*0201 in comparison to HLA DQA1*0201/DQB1*0201. *Hum. Immunol.* *53*, 195–205.
- Li, W., Moore, M.J., Vasilieva, N., Sui, J., Wong, S.K., Berne, M.A., Somasundaran, M., Sullivan, J.L., Luzuriaga, K., Greenough, T.C., et al. (2003). Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* *426*, 450–454.
- Li, F., Li, W., Farzan, M., and Harrison, S.C. (2005). Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* *309*, 1864–1868.
- Lippolis, J.D., White, F.M., Marto, J.A., Luckey, C.J., Bullock, T.N., Shabanowitz, J., Hunt, D.F., and Engelhard, V.H. (2002). Analysis of MHC class II antigen processing by quantitation of peptides that constitute nested sets. *J. Immunol.* *169*, 5089–5097.
- Liu, S., Xiao, G., Chen, Y., He, Y., Niu, J., Escalante, C.R., Xiong, H., Farmar, J., Debnath, A.K., Tien, P., and Jiang, S. (2004). Interaction between heptad repeat 1 and 2 regions in spike protein of SARS-associated coronavirus: implications for virus fusogenic mechanism and identification of fusion inhibitors. *Lancet* *363*, 938–947.
- Lorente, E., Martín-Galiano, A.J., Barnea, E., Barriga, A., Palomo, C., García-Ariaza, J., Mir, C., Lauzurica, P., Esteban, M., Admon, A., and López, D. (2019). Proteomics Analysis Reveals That Structural Proteins of the Virus Core and Involved in Gene Expression Are the Main Source for HLA Class II Ligands in Vaccinia Virus-Infected Cells. *J. Proteome Res.* *18*, 900–911.
- Lorente, E., Barnea, E., Mir, C., Admon, A., and López, D. (2020). The HLA-DP peptide repertoire from human respiratory syncytial virus is focused on major structural proteins with the exception of the viral polymerase. *J. Proteomics* *221*, 103759.
- Malaker, S.A., Ferracane, M.J., Depontieu, F.R., Zarling, A.L., Shabanowitz, J., Bai, D.L., Topalian, S.L., Engelhard, V.H., and Hunt, D.F. (2017). Identification and Characterization of Complex Glycosylated Peptides Presented by the MHC Class II Processing Pathway in Melanoma. *J. Proteome Res.* *16*, 228–237.
- Mazanetz, M.P., Marmon, R.J., Reisser, C.B., and Morao, I. (2012). Drug discovery applications for KNIME: an open source data mining platform. *Curr. Top. Med. Chem.* *12*, 1965–1979.
- Mellins, E., Woelfel, M., and Pious, D. (1987). Importance of HLA-DQ and -DP restriction elements in T-cell responses to soluble antigens: mutational analysis. *Hum. Immunol.* *18*, 211–223.
- Meunier, S., Hamze, M., Karle, A., de Bourayne, M., Gdoura, A., Spindelreher, S., and Maillere, B. (2019). Impact of human sequences in variable domains of therapeutic antibodies on the location of CD4 T-cell epitopes. *Cell Mol. Immunol.* *17*, 656–658.
- Millet, J.K., and Whittaker, G.R. (2015). Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. *Virus Res.* *202*, 120–134.
- Ovsyannikova, I.G., Johnson, K.L., Naylor, S., Muddiman, D.C., and Poland, G.A. (2003). Naturally processed measles virus peptide eluted from class II HLA-DRB1*03 recognized by T lymphocytes from human blood. *Virology* *312*, 495–506.
- Paul, S., Lindestam Arlehamn, C.S., Scriba, T.J., Dillon, M.B., Oseroff, C., Hinz, D., McKinney, D.M., Carrasco Pro, S., Sidney, J., Peters, B., and Sette, A. (2015). Development and validation of a broad scheme for prediction of HLA class II restricted T cell epitopes. *J. Immunol. Methods* *422*, 28–34.
- Paul, S., Sidney, J., Sette, A., and Peters, B. (2016). TepiTool: A Pipeline for Computational Prediction of T Cell Epitope Candidates. *Curr. Protoc. Immunol.* *114*, 18.19.11–18.19.24.
- Quarumby, V., Phung, Q.T., and Lill, J.R. (2018). MAPPs for the identification of immunogenic hotspots of biotherapeutics: an overview of the technology and its application to the biopharmaceutical arena. *Expert Rev. Proteomics* *15*, 733–748.
- Quinlan, B.D., Mou, H., Zhang, L., Guo, Y., He, W., Ojha, A., Parcels, M.S., Luo, G., Li, W., Zhong, G., et al. (2020). The SARS-CoV-2 receptor-binding domain elicits a potent neutralizing response without antibody-dependent enhancement. *bioRxiv*. <https://doi.org/10.1101/2020.2004.2010.036418>.
- Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). 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.* *48* (W1), W449–W454.
- Röhn, T.A., Reitz, A., Paschen, A., Nguyen, X.D., Schadendorf, D., Vogt, A.B., and Kropshofer, H. (2005). A novel strategy for the discovery of MHC class II-restricted tumor antigens: identification of a melanotransferrin helper T-cell epitope. *Cancer Res.* *65*, 10068–10078.
- Sallusto, F., and Lanzavecchia, A. (1994). Efficient presentation of soluble antigen by cultured human dendritic cells is maintained by granulocyte/macrophage colony-stimulating factor plus interleukin 4 and downregulated by tumor necrosis factor alpha. *J. Exp. Med.* *179*, 1109–1118.
- Schulze zur Wiesch, J., Lauer, G.M., Day, C.L., Kim, A.Y., Ouchi, K., Duncan, J.E., Wurcel, A.G., Timm, J., Jones, A.M., Mothe, B., et al. (2005). Broad repertoire of the CD4+ Th cell response in spontaneously controlled hepatitis C virus infection includes dominant and highly promiscuous epitopes. *J. Immunol.* *175*, 3603–3613.
- Sekiguchi, N., Kubo, C., Takahashi, A., Muraoka, K., Takeiri, A., Ito, S., Yano, M., Mimoto, F., Maeda, A., Iwayanagi, Y., et al. (2018). MHC-associated peptide proteomics enabling highly sensitive detection of immunogenic sequences for the development of therapeutic antibodies with low immunogenicity. *MAbs* *10*, 1168–1181.
- Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., and Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proc. Natl. Acad. Sci. USA* *117*, 11727–11734.
- Sidney, J., Southwood, S., Moore, C., Oseroff, C., Pinilla, C., Grey, H.M., and Sette, A. (2013). Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. *Curr. Protoc. Immunol.* *Chapter 18*, Unit 18.3.
- Speir, J.A., Abdel-Motal, U.M., Jondal, M., and Wilson, I.A. (1999). Crystal structure of an MHC class I presented glycopeptide that generates carbohydrate-specific CTL. *Immunity* *10*, 51–61.
- Strug, I., Calvo-Calle, J.M., Green, K.M., Cruz, J., Ennis, F.A., Evans, J.E., and Stern, L.J. (2008). Vaccinia peptides eluted from HLA-DR1 isolated from virus-infected cells are recognized by CD4+ T cells from a vaccinated donor. *J. Proteome Res.* *7*, 2703–2711.
- Tseng, C.T., Sbrana, E., Iwata-Yoshikawa, N., Newman, P.C., Garron, T., Atmar, R.L., Peters, C.J., and Couch, R.B. (2012). Immunization with SARS coronavirus vaccines leads to pulmonary immunopathology on challenge with the SARS virus. *PLoS ONE* *7*, e35421.

Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., and Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* *47* (D1), D339–D343.

Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., and Veerler, D. (2020). Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* *181*, 281–292.e6.

Walsh, R.E., Lannan, M., Wen, Y., Wang, X., Moreland, C.A., Willency, J., Knierman, M.D., Spindler, L., Liu, L., Zeng, W., et al. (2020). *Post-hoc* assessment of the immunogenicity of three antibodies reveals distinct immune stimulatory mechanisms. *MAbs* *12*, 1764829.

Wan, Y., Shang, J., Sun, S., Tai, W., Chen, J., Geng, Q., He, L., Chen, Y., Wu, J., Shi, Z., et al. (2020). Molecular Mechanism for Antibody-Dependent Enhancement of Coronavirus Entry. *J. Virol.* *94*, e02015-19.

Wantuch, P.L., Sun, L., LoPilato, R.K., Mousa, J.J., Haltiwanger, R.S., and Avci, F.Y. (2020). Isolation and characterization of new human carrier peptides from two important vaccine immunogens. *Vaccine* *38*, 2315–2325.

Yang, J., James, E., Roti, M., Huston, L., Gebe, J.A., and Kwok, W.W. (2009). Searching immunodominant epitopes prior to epidemic: HLA class II-restricted SARS-CoV spike protein epitopes in unexposed individuals. *Int. Immunol.* *21*, 63–71.

Zhou, H., Fang, Y., Xu, T., Ni, W.J., Shen, A.Z., and Meng, X.M. (2020a). Potential Therapeutic Targets and Promising Drugs for Combating SARS-CoV-2. *Br. J. Pharmacol.* *177*, 3147–3161.

Zhou, P., Yang, X.-L., Wang, X.-G., Hu, B., Zhang, L., Zhang, W., Si, H.-R., Zhu, Y., Li, B., Huang, C.-L., et al. (2020b). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* *579*, 270–273.

STAR★METHODS

KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| Anti-pan MHC-II (Tu39) | In-house | RRID:AB_393926 |
| Biological Samples | | |
| Frozen PBMCs | Discovery Life Sciences | http://www.dls.com/biopharma/cell-therapy-products/immune-cell-isolations/ |
| Chemicals, Peptides, and Recombinant Proteins | | |
| SARS-CoV-2 Spike protein (S-ECD) mutated furin site RRAR to RRAA | Genscript | Lot P9FD001 |
| Deposited Data | | |
| Raw mass spectrometry files | MASSIVE data repository (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) | ftp://massive.ucsd.edu/MSV000085456/ ; https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000085456 |
| Mass spectrometry peaks lists | MASSIVE data repository (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) | ftp://massive.ucsd.edu/MSV000085456/ ; https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000085456 |
| FASTA sequence database | MASSIVE data repository (https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp) | ftp://massive.ucsd.edu/MSV000085456/ ; https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?accession=MSV000085456 |
| Table S2. HLA-II Peptide Alignment to SARS-CoV-2 Extracellular Domain | Mendeley Data | https://data.mendeley.com/datasets/dk7zstnxsp/1 |
| Software and Algorithms | | |
| XTandem | The Global Proteome Machine Organization | https://www.thegpm.org/TANDEM/ |
| OMSSA | Geer LY, Wenger CD | https://github.com/coongroup/Compass |
| KNIME | https://www.knime.com/ | RRID:SCR_006164 |
| IEDB | Vita et al., 2019 | https://www.iedb.org |

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Robert Siegel (siegel_robert@lilly.com).

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Raw mass spectrometer data files, MGF files, and the search database are deposited in the MASSIVE data repository (<ftp://massive.ucsd.edu/MSV000085456/>; ID:MSV000085456, [<https://doi.org/10.25345/C51M7P>])

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Frozen PBMCs were obtained from 9 informed consent healthy donors (Discovery Life Sciences) according to local ethical practice.

| Donor ID | Donor Number | Age | Gender |
|----------|--------------|-----|--------|
| A | 39290 | 33 | Male |
| B | 39599 | 40 | Female |
| C | 39626 | 29 | Female |
| D | 39653 | 35 | Male |
| E | 40127 | 31 | Female |
| F | 40146 | 41 | Female |
| G | 40606 | 52 | Male |
| H | 40817 | 21 | Male |
| I | 42632 | 57 | Female |

METHOD DETAILS

Dendritic Cell Culture

Protocol was adapted as described (Sallusto and Lanzavecchia, 1994) with the following modifications. PBMCs were selected from the available inventory to have the broadest HLA-DRB1 diversity as possible. On day 0, frozen PBMCs were thawed in a 37°C water bath and washed twice with AIM-V media (Life Technologies, cat# 12055-083) with 1:20 diluted CTL Anti-Aggregate wash 20x solution (CTL, cat#CTL-AA-005). CD14+ cells were separated from the PBMCs by use of an AutoMACS instrument with anti-CD14 magnetic beads (Miltenyi Biotech, cat#130-050-201). After viability determination with Trypan Blue (Thermo Fisher Scientific, Countess cell counter) purified CD14+ mononuclear cells were resuspended in complete RPMI Media (RPMI 1640 with 2 mM L-glutamine (GIBCO, cat#11875-093); 5% Serum Replacement (Thermo Fisher Scientific, cat#A2596101)), 5mM HEPES (GIBCO, cat#15630-080), 1% of 100X MEM nonessential amino acids (GIBCO, cat#11140-050), 1% of 10,000 U/mL Penicillin/Streptomycin (HyClone, cat#SV30010), 1mM sodium pyruvate (GIBCO, cat#11360-070), 50 uM β -mercaptoethanol (Fisher Chemical, cat#O34461-100), and 3.5% of DMEM high glucose (GIBCO, cat#31053-028) containing 40ng/ml granulocyte monocyte-colony stimulating factor (human GM-CSF; Sargramostim, Sanofi-Aventis, NDC code 0024-5843-05) and 20 ng/mL IL-4 (R&D Systems, cat # 204-IL) to a density of 1×10^6 cells/ml and were differentiated into mature DCs in 6 well cell culture dishes (5 mL final culture volume) at 37°C and 5% CO₂. On day 4, immature DCs were loaded with 2.5 mg of recombinant HIS-FLAG tagged extracellular domain SARS-CoV-2 Spike protein from the SARS-CoV-2 Wuhan-Hu-1 strain (GenBank: YP_009724390) with the furin site mutated from RRAR to RRAA (Genescript) and were incubated for 24 hours. One donor's cells displayed a different morphology compared to the others donor's cells. DCs were matured on day 5 by adding lipopolysaccharide (LPS, 1 μ g/mL final; Sigma-Aldrich, cat#L5886) to the media.

HLA-II Peptide Isolation

The cells were lysed on Day 6 with RIPA lysis and extraction buffer (Thermo Fisher Scientific, cat # 89900, 25 mM Tris·HCl pH 7.6, 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate, 0.1% SDS) containing 1:1000 of 10 unit/uL DNase (Roche, cat# 04716728001) and 1 tablet of EDTA free protease inhibitors (Roche, Cat# 11836170001) per 10 mL of lysis buffer. The lysates were frozen at -80°C . An Agilent AssayMAP robot was used to isolate the HLA-II molecules in the lysate. One hundred microliters of 1 mg/mL biotinylated anti-pan HLA class II antibody (in house produced Tu39 clone) was immobilized on streptavidin cartridges (Agilent, SA-W, 5 uL) by passing over the cartridge at 5 uL/min. The cartridge was washed with 50 uL of PBS three times. Lysates were thawed, passed over a 0.45 μ m filter, and 1 mL of each sample was loaded onto a 96 well polypropylene plate. The lysate was aspirated into the syringes and the antibody loaded cartridge is attached to the syringe tip. The lysate is passed over the affinity cartridge at 5 uL/min at room temperature for 200 minutes. The cartridge is washed 2 \times 50 uL with 100mM ammonium acetate at 25 uL/min and once with 50 uL water at 25 uL/min. The cartridge is eluted with 50 uL of 5% acetic acid with 0.1% TFA at 2 uL/min into a 96 well polypropylene PCR plate. The eluted peptides were passed over a 10k MWCO spin filter treated with 1 mg/mL BSA (Sigma, 05470) and 100ug/mL angiotensin I peptide and washed with 5% acetic acid. The filtered material was loaded in a 96 well polypropylene PCR plate for mass spec analysis.

LC/MS analysis of HLA-II Eluted Peptides

The samples were analyzed with a Thermo LUMOS mass spectrometer using a Thermo easy 1200 nLC-HPLC system. The separation was carried out with a 75 μ m \times 7 cm YMC-ODS C18 column (New Objectives) coupled to a custom nanospray interface with an electrospray potential of 1.2 kV. The solvents were A - 0.1% formic acid in water (Thermo Fisher Scientific, Optima LC/MS Grade) and

B –80% acetonitrile with 0.1% formic acid (Thermo Fisher Scientific, Optima LC/MS Grade). The gradient was 65 minutes using a flow rate of 250 nL/min, starting with a 60 min 2%–55%B ramp followed by a 1 min 55%–100%B ramp and a 4 min hold at 100%B). The Lumos was run with a full scan at 240,000 resolution in the orbitrap followed by a 3 s data dependent MS/MS cycle comprised of ion trap rapid scans where +2 ions were fragmented by HCD(CE of 15,22,28) and +3 and +4 ions were fragmented by HCD(CE of 15,22,28) and ETHcD (Calibrated Charge-Dependent ETD parameters and supplemental HCD (CE of 50)).

QUANTIFICATION AND STATISTICAL ANALYSIS

The data were analyzed with the Lilly proteomics pipeline (Higgs et al., 2008). The data conditioning steps consisted of extraction from the vendor format, fitting parent ions for data dependent scans to theoretical isotope patterns and correcting the monoisotopic mass and charge of the parent ion, determining the fit of the parent ion isotope to the theoretical isotope pattern and filtering out ms/ms scans if the parent ions did not match the isotope pattern with a score of 0.6 or greater. From the filtered scans, an MGF file was created along with a table of spectral features for each spectrum.

The spectral identifications were performed with X! Tandem version 2017 and OMSSA version 2.1.7 search engines. A database was used consisting of the SARS-CoV-2 spike extracellular domain HIS-FLAG tagged protein and 2134 common human and bovine proteins identified from HLA-II bound peptides seen from Raji cells, DCs, and bovine proteins in the cell media.

The search engine parameters included a no enzyme search with a maximum missed cleavage site setting of 30, 10 ppm tolerance for parent ions, and 0.5 m/z tolerance for the fragment ions. Potential amino acid modifications included: Cysteine mods of free SH; disulfide; mercaptoethanolation; mono, di, and tri oxidation; and cysteinylolation; deamidation of glutamine and asparagine; methionine oxidation; tryptophan oxidation, deoxidation, oxidation to kynurenin. HCD spectra were searched for b- and y-ions and ETHcD were searched for c-, z-, b-, and y-ions.

False positive identifications were controlled by running the searches against a reversed version of the protein database and estimating false discovery rates. An iterative random forest classifier was trained using search results and spectral features to increase identification sensitivity in a manner similar to the percolator algorithm (Käll et al., 2007). The search results from X! Tandem and OMSSA were pooled and peptides with q-values < 0.20 were assigned to the smallest group of proteins that account for all identified peptides.

Pepnovo plus (Frank and Pevzner, 2005) was run on all the spectra with a parent ion tolerance of 10 ppm and a fragment tolerance of 0.5 Th. Modifications included were methionine oxidation, cysteinylolation, and disulfide formation. The output was filtered for peptide tags that matched the SARS-CoV-2 spike extra cellular domain HIS FLAG protein. Tag hits were checked against the results from the database search. All tag hits were matched to the database search results indicating that there were no unknown modifications present in the results.

The pipeline output was analyzed using KNIME 3.3 (Mazanetz et al., 2012) to merge all the donor search results, manual review of the ms/ms spectra of the identifications to confirm presence of at least 4 contiguous fragment ions to matched peptide sequence, align the peptides to the SARS-CoV-2 spike extracellular domain HIS FLAG protein, and create an excel file with the alignment.

The identified peptides from the SARS-CoV-2 spike glycoprotein were clustered with the IEDB Epitope Cluster Analysis tool v1.0 using the default settings. The clustering was manually adjusted to group a continuous amino acid run of at least 9 residues after sorting by donor. This adjustment was reflected in the cluster number as a decimal point after the assigned cluster number from the IEDB algorithm. Predicted SARS-Cov-2 spike protein peptides (Grifoni et al., 2020a) were matched with the MAPPs cluster if they shared at least a 9 amino acid overlap to the minimum cluster sequence.

The identified HLA-II peptides were checked for exact matches using the UNIX grep command. The sequences searched were the human database (UniProt version 2020_02), spike proteins from SARS-CoV UniProt: P59594), human coronavirus NL63 (UniProt: A0A5B9BGI8), 229E (UniProt: A0A223FUI6), HKU1 (UniProt: U3NAI2), OC43 (UniProt: P36334), and MERS (GenBank: YP_009047204.1). An imperfect match search was run using the UNIX command agrep to look for 1 or 2 mismatched amino acids. The searches were done for each peptide with insertion and deletions scores set to 100 to discount any insertion or deletions during the search. The imperfect match search was run against the human database (UniProt version 2020_02).