Identification of presented SARS-CoV-2 HLA class I and HLA class II peptides using HLA peptidomics

Graphical abstract

Highlights
- HLA peptidomics enable the identification of SARS-CoV-2-derived HLA peptides
- SARS-CoV-2 peptides can be derived from canonical and non-canonical ORFs
- Shared SARS-CoV-2 peptides are identified in different cell types
- Several SARS-CoV-2 peptides are immunogenic

In brief
Using HLA peptidomics, Nagler et al. identify SARS-CoV-2-derived peptides presented by highly prevalent HLA-I and HLA-II. Identified peptides are derived from various canonical and non-canonical viral ORFs, are shared between different cells, and are immunogenic. As the identified peptides are unique to SARS-CoV-2, they could potentially serve as future treatment targets.

Authors
Adi Nagler, Shelly Kalaora, Chaya Barbolin, ..., Nir Paran, Pia Kvistborg, Yardena Samuels

Correspondence
yardena.samuels@weizmann.ac.il

Nagler et al., 2021, Cell Reports 35, 109305
June 29, 2021 © 2021 The Author(s).
https://doi.org/10.1016/j.celrep.2021.109305
Identification of presented SARS-CoV-2 HLA class I and HLA class II peptides using HLA peptidomics

Adi Nagler,1,8 Shelly Kalaora,1 Chaya Barbolin,1 Anastasia Gangaev,2 Steven L.C. Ketelaars,2 Michal Alon,1 Joy Pai,5 Gil Benedek,3 Yfat Yahalom-Ronen,4 Noam Erez,4 Polina Greenberg,1 Gal Yagel,1 Aviyah Peri,1 Yishai Levin,6 Ansuman T. Satpathy,5 Erez Bar-Haim,3 Nir Paran,4 Pia Kvistborg,3 and Yardena Samuels1,9,*
1Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel
2Division of Molecular Oncology and Immunology, the Netherlands Cancer Institute, the Netherlands
3Tissue Typing and Immunogenetics Unit, Hadassah Medical Organization and Faculty of Medicine, Hebrew University of Jerusalem, Israel
4Department of Infectious Diseases, Israel Institute for Biological Research, Ness Ziona, Israel
5Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA
6The de Botton Institute for Protein Profiling, The Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel
7Department of Biochemistry and Molecular Genetics, Israel Institute for Biological Research, Ness Ziona, Israel
8These authors contributed equally
9Lead contact
*Correspondence: yardena.samuels@weizmann.ac.il
https://doi.org/10.1016/j.celrep.2021.109305

SUMMARY

The human leukocyte antigen (HLA)-bound viral antigens serve as an immunological signature that can be selectively recognized by T cells. As viruses evolve by acquiring mutations, it is essential to identify a range of presented viral antigens. Using HLA peptidomics, we are able to identify severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-derived peptides presented by highly prevalent HLA class I (HLA-I) molecules by using infected cells as well as overexpression of SARS-CoV-2 genes. We find 26 HLA-I peptides and 36 HLA class II (HLA-II) peptides. Among the identified peptides, some are shared between different cells and some are derived from out-of-frame open reading frames (ORFs). Seven of these peptides were previously shown to be immunogenic, and we identify two additional immunoreactive peptides by using HLA multimer staining. These results may aid the development of the next generation of SARS-CoV-2 vaccines based on presented viral-specific antigens that span several of the viral genes.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) caused a global pandemic of coronavirus disease 2019 (COVID-19) (Zhu et al., 2020). Effective protective vaccines against COVID-19 symptomatic disease that target the spike protein have been developed (Jackson et al., 2020; Mulligan et al., 2021). Although most treatment development efforts focus on creating anti-viral antibody responses, T cell immunity also appears to play a key role in the immune response during COVID-19 (Dan et al., 2021; Rodda et al., 2021). In contrast to the decrease in stable spike- and nucleocapsid-specific antibody responses, functional T cell responses remain robust up to 6 months post-infection (Bilich et al., 2021).

Viral antigens are proteolytically processed by infected cells into peptides and then bound to human leukocyte antigen (HLA) molecules to be presented to T cells (Croft et al., 2019; Madden et al., 1993). This “HLA peptidome” serves as an immunological signature that can be selectively recognized by CD8+ and CD4+ T cells by their T cell receptor (TCR), potentially leading to cell lysis and catalyzing other immune responses (Welters et al., 2003; Yerly et al., 2008; Zhao et al., 2002). Compared to B cell responses that recognize external viral epitopes, T cell responses can be potentially mediated by all viral proteins. As a result, apart from the spike protein, other SARS-CoV-2 proteins have the potential to serve as T cell antigens. Bioinformatics predictions for HLA class I (HLA-I) and HLA class II (HLA-II) binding affinity have been applied to detect SARS-CoV-2 viral epitopes. These epitopes could be recognized by T cells and further generate sustainable memory populations (Campbell et al., 2020; Grifoni et al., 2020; Joshi et al., 2020; Kiyotani et al., 2020; Lin et al., 2020; Nelde et al., 2020; Nguyen et al., 2020; Poran et al., 2020). The current knowledge of SARS-CoV-2 peptides are based mainly on data generated by biochemical binding assays (IMMUNITRACK website) and reactivity assays (Grifoni et al., 2020; Nelde et al., 2020; Poran et al., 2020; Sekine et al., 2020; Woldemeskel et al., 2020; Zhang et al., 2020) of predicted peptides. Although these bioinformatics tools
can be extremely useful, there is no certainty whether all of the SARS-CoV-2 predicted peptides are indeed processed and presented by antigen-presenting cells (APCs) or infected cells during SARS-CoV-2 infection. To overcome this limitation, in this study, we combined bioinformatics tools with HLA peptidomics to identify the naturally processed and presented HLA-I and HLA-II peptides by using two experimental approaches (Figure 1). To identify HLA peptides that are potentially presented by a wide range of the human population, we selected the most frequent HLA alleles in the human population and used mono-allelic cells or multi-allelic cells that express them endogenously or by overexpression. The selected cells were transduced with viral genes or infected with SARS-CoV-2 virus followed by HLA peptidomics, leading to the identification of presented HLA-I and HLA-II viral antigens.

The discovery of non-canonical open reading frames (ORFs) in human and viral genomes has been of particular interest (Finkel et al., 2020b; Ingolia et al., 2009, 2011; Nomburg et al., 2020; Prensner et al., 2020; Stern-Ginossar et al., 2012), specifically in relation to the identification of presented HLA antigens derived from non-canonical ORFs by human tumors (Chen et al., 2020; Chong et al., 2020; Ingolia et al., 2014; Ouspenskaia et al., 2020; Starck and Shastri, 2016). By combining the canonical and non-canonical translational landscape of SARS-CoV-2 (Finkel et al., 2020a) with HLA peptidomics, we were able to identify viral peptides from both the canonical and non-canonical ORFs.

Furthermore, we have identified shared peptides that were presented by several different cell lines, strengthening their identification and ability to be processed and presented by diverse types of cells. Importantly, our study describes 62 SARS-CoV-2 peptides presented from different viral genes and reading frames.

**RESULTS**

**Identification of SARS-CoV-2 HLA peptides by using overexpression of SARS-CoV-2 genes**

To assess the presentation of viral peptides both by endogenous and overexpressed HLA-I molecules, we used 721.221 B cells with mono-allelic expression of the most frequent HLA-I alleles or IHW01070 and IHW01161 B cells and Calu-6 (lung adenocarcinoma) cell line with endogenous multi-allelic expression of the most frequent HLA-I alleles. Both of these cell systems co-expressed specific viral genes. We chose to use B cells because they not only are infected by viruses (Gu et al., 2005; Pontelli et al., 2020; Sorem et al., 2009; Spear and Longnecker, 2003) but also play a central role in presenting the peptides to the immune system (Cheng et al., 1999; Hong et al., 2018) and lung-derived cell line, due to its relevance in the disease.

To better identify virus-presented peptides that are shared across patients, we focused on HLA alleles that were present in a large fraction of the world population. To this end, we first selected the most frequent HLA-I alleles according to the allele frequency net database (http://www.allelefrequencies.net/; Figure S1A). This selection resulted in the inclusion of at least one of the most frequent HLA-A/B/C alleles for each population (Figure S1B).

We overexpressed the SARS-CoV-2 envelope, membrane, nucleocapsid, and nonstructural protein 6 (nsp6) genes in the B cells outlined above. We specifically focused on these four genes, as previous studies, by using pools of predicted peptides derived from these four SARS-CoV-2 genes, showed a high frequency of reactive T cells from exposed and unexposed individuals (Grifoni et al., 2020; Sekine et al., 2020). To assess in which HLA-I context each SARS-CoV-2 gene should be
overexpressed, we performed HLA binding predictions of the viral genes to the most frequent HLA-I alleles and selected alleles with the highest combination of allele frequency and number of predicted alleles for each gene (Figures S1C and S1D).

We then used HLA peptidomics to profile HLA-I and HLA-II-bound antigens as previously described (Abelin et al., 2017, 2019; Bassani-Stemberg et al., 2016; Kalaora et al., 2016, 2018, 2021). All HLA peptidomics analyses were performed in triplicate. The raw data from each HLA peptidomics analysis was searched using MaxQuant software against the relevant overexpressed viral gene and the human proteome in the same analysis. This analysis revealed 15 unique HLA-I and 36 unique HLA-II-associated peptides, respectively (Table S1). The length distribution of the identified peptides was consistent with the expected length of HLA-I and HLA-II peptides (Figure S2A).

The clustering of the 8–13 identified amino acid HLA-I peptides from each cell line showed a reduced amino acid complexity of the peptides, as expected for HLA-I peptides, and matched the patient’s HLA allele motif (see STAR Methods section). We arbitrarily chose 28 of the peptides for validation by comparing their tandem mass spectrometry (MS/MS) spectra to that of synthetic peptides; all peptides were validated (see STAR Methods section). Similarly, we spikes stable isotopically labeled synthetic peptides that co-eluted with the endogenous peptides, further validating their identification (see STAR Methods section). All the identified viral peptides showed a correlation between their retention time and predicted hydrophobicity, supporting their identification (see STAR Methods section).

Identification of SARS-CoV-2-derived peptides in virus-infected cells

To complement our overexpression system, we infected the HEK293T cell line stably expressing angiotensin-converting enzyme 2 (ACE2) and IHW01070, Calu-3 cell lines with SARS-CoV-2. First, we confirmed that these cells express the ACE2 receptor and transmembrane protease serine 2 (TMPRSS), which are known to be required for SARS-CoV-2 entry into the cells (Hoffmann et al., 2020; Matsuyama et al., 2020; Wan et al., 2020; Zhou et al., 2020; Figure S2B). Infection conditions were optimized to allow a high number of infected cells while preserving their viability (Figure S2C), and samples were processed for HLA-I peptidomics.

We assessed whether any SARS-CoV-2 peptides in the infected cells may have been derived from non-canonical ORFs by using recently published novel ORFs (Finkel et al., 2020a). Indeed, we were able to identify 2 HLA-I peptides derived from internal out-of-frame ORFs present in the coding region of spike and nucleocapsid proteins in HEK293T cells. Peptide GPMVLRGLIT was identified from ORF 5′orf1′/2 and was predicted to bind the B*07:02 allele; SLEDKAFOQL was identified from ORF9b (an internal out-of-frame ORF in the coding region of nucleocapsid) and was predicted to bind the A*02:01 and C*07:02 alleles.

We also identified 10 different HLA-I peptides derived from canonical ORFs of the virus. Two peptides were derived from the spike protein, as follows: NEVAKNLNESL was identified in IHW01070 cells and matched their B*40:01 allele and TGSNVFQCTR was identified in the Calu-3 cells and matched their A*68:01 allele. Two peptides were identified in the Calu-3 cells and derived from nsp3, namely, STTNTIVTR and YYTSNPTTF, and they matched the A*68:01 and A*24:02 alleles, respectively. In Calu-3 cells, FTIGTVTLK was derived from ORF5a and matched the A*68:01 allele, and HSSGVTREL was derived from nsp1 and matched the C*15:02 allele. Four peptides were derived from nucleocapsid; one of them, APRITFGGP, was identified in both HEK293T and Calu-3 cells (matching the B*07:02 allele), and three peptides overlapped to this peptide (RITFGGPSD-A*03:01 was identified in HEK293T cells and NAP-RITFGGP-A*68:01 and ITFGPSDSTGSQNGER-A*68:01 in Calu-3 cells) (Table S1). SARS-CoV-2 -derived peptides were not identified in uninfected cells that underwent HLA peptidomics. Thus, our results further substantiate the validity of our identified viral peptides.

To assess whether the infection with SARS-CoV-2 alters the levels of human HLA-I peptides, we compared the intensity of peptide presentation between infected and non-infected IHW01070 cells (two-sided Student’s t test, permutation-based false discovery rate [FDR] = 0.05, S0 = 1; Figure 2A). Among the peptides that were more highly presented after infection, we observed a statistically significant enrichment of immune-related pathways, including the JAK/STAT signaling, FAS signaling, and B cell and T cell activation (Fisher’s exact test, FDR < 0.05; Figure 2B).

We performed proteomics analysis from the flow through of the HLA peptidomics lysates to compare infected and non-infected IHW01070 cells and identify the pathways that were altered due to the infection (two-sided Student’s t test, permutation-based FDR = 0.05, S0 = 0.1; Figure 2C). This analysis revealed a significant downregulation in proteins involved in type I interferon (IFN) response as previously shown (Bost et al., 2020). Interestingly, the IFNγ-induced immunoproteasome subunits (PSMB9 and PSMB10) (Kloetzel, 2001) were downregulated in the infected cells, suggesting that the virus affects the presented peptidome by altering the proteasome cleavage specificities.

As we observed a downregulation of proteins involved in the IFN signaling pathway in the proteomic analysis of the cells, and we observed an increase in presentation of HLA peptides from this pathway (e.g., JAK1, JAK2, STAT1, STAT3, and STAT6), it may be suggested that degradation of the proteins in this pathway is a result of the viral infection.

Identification of shared SARS-CoV-2-derived HLA peptides

We found three HLA-I peptides that were shared between samples. The first peptide is NSSPDGDIYGY, which was derived from nucleocapsid and matched the A*01:01 allele. It was identified in the IHW01070, IHW01161, and Calu-6 cells that endogenously expressed A*01:01, as well as in 721.221 cells that overexpressed both A*01:01 and nucleocapsid. The second shared peptide APRITFGGP was derived from nucleocapsid and identified in both SARS-CoV-2-infected HEK293T and Calu-3 cells as well as in 721.221 cells with overexpression of the B*07:02 allele and nucleocapsid. Finally, peptide FLPLSLATV was derived from nsp6 and identified in IHW01070 cells as well as in 721.221 cells with overexpression of the A*02:01 allele and nsp6.
We further identified overlapping peptides derived from the same protein, which were presented on both HLA-I and HLA-II. Specifically, a HLA-II nested set of 25 peptides, which matched the DRB1*01:02 allele, was identified in 721.221 cells overexpressing the membrane gene. Furthermore, nine nested HLA-II peptides from nucleocapsid that matched the DRB1*01:02 allele were identified in 721.221 cells overexpressing the nucleocapsid gene. The HLA-I (ATEGALNTPK) and HLA-II (KDGIIWVATEGALN) peptides from the nucleocapsid protein showed a 7-amino-acid overlap and were identified in 721.221 overexpressing the A11:01 allele and in IHW01070, both overexpressing nucleocapsid; and four overlapping HLA-I from nucleocapsid were identified in SARS-CoV-2-infected cells, as discussed above (APRITFGGP, RITFGGPSD, NAPRITFGGP, and ITFGGPSDTGNSQNGER). Two HLA-I peptides from the nsp6 protein, namely, DYLVSTQEF and VYDYLVSTQEF, were identified in 721.221 cells that overexpressed nsp6 and the A*24:02 allele (Figure 3).

A comparison of our HLA-peptidomics-derived peptides to our matrix of predicted peptides of the same gene and allele showed weak and strong binding prediction for the identified peptides, indicating that peptides predicted to be weak binders can also be presented.

Importantly, five of our identified HLA-I peptides were previously shown, by using peptide stability and affinity assays, to bind the same allele of the cells in which we identified the peptide, supporting our findings (IMMUNITRACK website; Covid19 Intavis_Immunitrack stability dataset 1; Cheung et al., 2007; Sylvester-Hvid et al., 2004; Table S1).

**Figure 2. Differently presented peptide repertoire in IHW1070 B cell line after infection with SARS-CoV-2**

(A) A volcano plot was used to identify the peptides that were differentially presented by the cell’s HLA-I molecules of infected compared to the uninfected control. HLA peptidomics experiments were done in three biological replicates. The red dots indicate proteins involved in immune regulation pathways, indicated in (B).

(B) The table indicates the pathways that were found to be significantly enriched of the peptides that were significantly more presented after SARS-CoV-2 infection.

(C) Volcano plot of proteins identified in the proteomic analysis of the cells, comparing infected and non-infected IHW01070. Proteomic experiments were done in three biological replicates. Type I interferon response proteins are marked in red, and beta proteasome subunits are marked in blue.

We assessed the similarity of identified SARS-CoV-2 peptides to the proteome of other coronavirus family members to which one might have prior immunity. Among predicted SARS-CoV-2 peptides, an identical sequence or peptides with one amino acid
discovery were identified only in nucleocapsid, membrane, envelope, and spike proteins (Figure 4). Some of the predicted peptides overlap with SNPs found in the different strains of the SARS-CoV-2 genome (Figure 4). From the presented peptides identified in our study, 10 peptides such as GMSRIGMEV derived from the nucleocapsid showed 100% similarity to the SARS-CoV-1 virus (Cheung et al., 2007; Ohno et al., 2009; Sylvester-Hvid et al., 2004; Tsao et al., 2006). Eighteen peptides showed similarity to SARS-CoV-1 with one amino acid substitution (Table S2). No similar peptides (exact sequence or one amino acid substitution) were found when comparing to other CoV species (HCoV-NL63, HCoV-229E, HCoV-OC43, HCoV-HKU1, and MERS-CoV). In addition, none of our identified peptides were found among the different strains of the SARS-CoV-2 SNPs, suggesting that these peptides could be used for patients infected with different SARS-CoV-2 strains.

We compared the amino acid sequence of the SARS-CoV-2 peptides that were predicted to bind the most frequent alleles to the human genome. We found 2 different peptides that had the same sequence and that 281/9,836 unique peptides (2.8%) were homologous in their sequence to human peptides (with one amino acid substitution, Figure 4; see STAR Methods section); some of them were similar to peptides from different human genes. Most of the similar human sequences were predicted to bind the same allele as the predicted viral peptide, increasing the chance of cross-reactivity to these human peptides in patients that previously had an immune response to the viral peptides. In contrast, when we assessed whether our identified presented SARS-CoV-2 peptides are unique in their sequence or whether they have homology to peptides in the human proteome, we found no amino acid similarities, even with a flexibility of one amino acid change.

**Presented SARS-CoV-2 peptides are immunogenic**

Several studies have reported T cell reactivity toward the SARS-CoV-2 HLA-predicted peptides (Grifoni et al., 2020; Nelde et al., 2020; Sekine et al., 2020; Woldemeskel et al., 2020; Zhang et al., 2020). We compiled a map of all the available data of experimentally validated reactive SARS-CoV-2 peptides (Le Bert et al., 2020; Minervina et al., 2020; Neide et al., 2020; Poran et al., 2020; Schulien et al., 2020; Shomuradova et al., 2020; Snyder et al., 2020; Tarke et al., 2020; Woldemeskel et al., 2020; Wong et al., 2020) and searched it to see if it includes any of our identified peptides (Figure 4). Indeed, seven of our peptides were previously found to be immunogenic in blood samples of COVID-19 patients. The immunodominant epitopes ATEGALNTPK and KTFFPTEPK were found in 9/11 (82%) and 7/11 (64%) tested samples, respectively (Nelde et al., 2020). We identified their presentation by conducting HLA peptidomics of 721.221 B cells overexpressing nucleocapsid and the A*11:01 or A*03:01 allele, respectively. Furthermore, we identified FVKHKKHAFL as a presented HLA-I peptide by the IHW01070 B cells overexpressing nsp6 and endogenously expressing the B*08:01 HLA allele, which is predicted to bind this peptide. Importantly, this peptide was previously found to be recognized by CD8+ T cells in 1/12 (8%) patients (Nelde et al., 2020). The presented HLA-II peptide LSYYKLQASPVRAGD was identified in the 721.221 B cells overexpressing membrane and expressing the matching DRB*01:02 allele endogenously. This HLA-II peptide was previously found to be recognized by CD4+ T cells in 10/12 patients (Nelde et al., 2020). The presented HLA-I peptide FTTGVTTLK was identified in Calu-3 cells infected with SARS-CoV-2 and found to bind to A*08:01. The HLA-I peptide VYM-PASWVM was identified in 721.221 B cells overexpressing nsp6 and the A*24:02. These peptides were previously found to be recognized by CD8+ T cells in 1/1 (100%) patients (Tarke et al., 2020). Finally, GMSRIGMEV, which was found to bind to the B*13:02 allele in this study, was previously identified to bind A*02:01 and to be immunogenic both in human blood samples and in an A*02:01 mouse model (Cheung et al., 2007; Ohno et al., 2009; Tsao et al., 2006).

To test for CD8+ T cell recognition, we selected two SARS-CoV-2 HLA-I peptides identified in this study that were not previously tested for reactivity and were identified in different samples. We used peripheral blood mononuclear cell (PBMC) samples from 6 COVID-19 patients with diverse disease severities and two healthy donors (Figure 5; Figure S3). CD8+ T cell recognition of the SARS-CoV-2 epitopes was assessed by multiplexing fluorescent peptide-HLA (pHLA) multimers. Our analysis revealed CD8+ T cell responses specific for the NSSPDQIYY epitope in two HLA-A*01:01-positive patients and for the FLLPSLATV epitope in two HLA-C*07:01-positive patients (Figure 5; Figure S3). The magnitude of the NSSPDQIYY-specific CD8 T cell responses was 0.017% and 0.025% of total CD8+ T cells in COVID-131 and COVID-007, respectively. The magnitude of the FLLPSLATV-specific CD8+ T cell responses was 0.590% and 0.067% of the total CD8+ T cells in patients COVID-131 and COVID-224, respectively. We observed a higher frequency of antigen-specific CD8+ T cells in COVID-19 patients with severe or critical disease than that of asymptomatic patients or healthy non-exposed individuals. Together with the findings from previous studies, our results confirm the recognition of identified epitopes by CD8+ T cells.

**DISCUSSION**

By using our HLA peptidomics approach on cells overexpressing SARS-CoV-2 genes and virally infected cells, we yielded several important findings. First, we identified 62 presented HLA peptides derived from 8 SARS-CoV-2 proteins. As we have identified both HLA-I- and HLA-II-bound epitopes, vaccines containing such epitopes will potentially lead to CD4+ T cell, CD8 + T cell, and B cell stimulation, driving both humoral and cellular immunity. In addition, targeting both structural and functional viral proteins, which have a crucial role in the viral life cycle, may help to create a broader immune response. Second, we have identified three shared presented peptides derived from two SARS-CoV-2

---

**Figure 3. SARS-CoV-2-derived shared HLA peptides and presentation hotspots**

Schematic representation of all peptides with identified SARS-CoV-2 gene overexpression and infections. Each cell line is marked by a different color dot, HLA-I peptides are marked in red box, and HLA-II peptides are marked in blue box. Peptides found to be immunoreactive are marked in red, and peptides shown to bind the corresponding HLA are marked in green.
genes. Third, in order to study the full landscape of the SARS-CoV-2 antigen repertoire, we included the non-canonical ORFs in our analysis, which led us to identify two viral peptides derived from out-of-frame-ORFs. Fourth, seven of the identified peptides were previously shown to be immunogenic, and we have shown the reactivity of two additional peptides, of which both are shared peptides, by using COVID-19-patient-derived T cells. Fifth, our presented HLA peptide sequences are not found in the human proteome, reducing the likelihood of immune-related adverse and pathological effects. And sixth, our presented HLA peptide sequences did not overlap with SNPs of other SARS-CoV-2 strains, making them relevant for targeting different strains of the virus. Despite these insights, we acknowledge the limitations of identifying presented viral peptides by using HLA peptidomics, as it relies on peptide intensity, which at times may be below the detection level, when viral proteins are lowly expressed.

Identification of SARS-CoV-2 antigens in previous studies was primarily carried out using approaches based on the prediction of virus-presented peptides and, in some cases, the assessment of their reactivity using HLA-matched T cells (Campbell et al., 2020; Grifoni et al., 2020; Joshi et al., 2020; Kiyotani et al., 2020; Lin et al., 2020; Nelde et al., 2020; Nguyen et al., 2020; Poran et al., 2020). However, although these approaches are extremely valuable, they have several drawbacks, as follows: (1) as the HLA-peptide-processing steps are not completely elucidated, they cannot be fully integrated into the prediction algorithm (Abelin et al., 2017, 2019; Bassani-Sternberg et al., 2016; Kalaora et al., 2016, 2018, 2021). (2) In the immune-reactivity assays, which are performed with blood samples from non-exposed individuals, there is no certainty that the antigen tested can actually be presented or that the results are due to cross-reactivity (of peptides that have sequence similarity to other viruses) (Jeyanathan et al., 2020; Mateus et al., 2020) or due to the fact that these are non-self antigens. Testing peptide reactivity by using blood from infected patients, in which a high frequency of T cells recognizing SARS-CoV-2 peptides is observed, gives more certainty that the peptide was presented by infected cells.

Our methodology is highly complementary to the prediction and T cell reactivity assessments currently being performed, as it allows a more comprehensive evaluation of which viral antigens may be presented by infected cells and APCs and elicit an immune response. Selecting HLA-presented areas of the viral proteome that are conserved between the different variants of the SARS-CoV-2 virus could help develop vaccines that take into account viral evolution and development of viral variants.

**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**
  - SARS-CoV-2 virus
  - Human samples
- **METHOD DETAILS**
  - Selection of most frequent HLA-I alleles in the world population
  - Cells lines
  - HLA typing
  - Pooled stable expression of HLA-I alleles in 721.221 cells
  - Pooled stable expression SARS-CoV-2 protein expressing B cells and Calu-6 cells
  - Quantitative PCR assay
  - Production and purification of membrane HLA molecules
  - Mass-spectrometry analysis of eluted HLA peptides
  - Identification of SARS-CoV-2-derived HLA peptides
  - Hydrophobicity index calculation
  - Validation using synthetic peptides
  - Gibbs clustering
  - Flow cytometry
  - SARS-CoV-2 viral infection
  - SARS-CoV-2 cell infection efficiency
  - Proteomic analysis of infected cells
  - Analysis of differentially expressed proteins and presented peptides
  - Pathway analysis
  - Similarity of SARS-CoV-2 peptides to other CoV species
  - Finding SNPs in the SARS-CoV-2 genome
  - Similarity of SARS-CoV-2 peptides to human proteome
  - Generation of fluorescent pHLA multimers
  - Surface staining with pHLA multimers and antibodies
  - Flow cytometer setting for pHLA multimers
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

**SUPPLEMENTAL INFORMATION**

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109305.

**ACKNOWLEDGMENTS**

We thank Prof. Nevan J. Krogan and Dr. David E. Gordon for providing SARS-CoV-2 envelope, nucleocapsid, nonstructural protein-6 and membrane pLVX-EF1alpha-IRES-Puro plasmids. We thank Prof. Moshe Kotler for his advice regarding viral pathogenesis. We thank Tali Wiesel for the graphical scheme.

---

**Figure 4. Schematic map of predicted and presented peptides of SARS-CoV-2**

predicted peptides are marked in black. Predicted peptides that overlapped with peptides derived from the CoV family are marked in green if they were identical or light green if similar with one substitution. Predicted peptides that were similar (with one substitution) to human peptides are marked in red. All peptides previously found to be immunogenic in different studies are marked in blue. Presented peptides, identified in this study, are marked in purple. The frequency of SNPs in the SARS-CoV-2 variants are represented in the line plot.
We thank Dr. Tamar Ziv, Shani Ben-Iru, Keren Bendalak and the Smoler Protein Research Center, Technion, Israel for the proteomic analysis of infected cells. This work was funded by the Intramural Research Programs of the National Cancer Institute. Y.S. is supported by the Israel Science Foundation grant No. 696/17, the European Research Council (ERC) under the European Union’s Horizon 2020 Programme for Research and Innovation (grant agreement No. 754282), ERC (CoG-770854), MRA (#662106), Rising Tide Foundation, Henry Chanock Krenter Institute for Biomedical Imaging and Genomics, Estate of Alice Schwarz-Gardos, Estate of John Hunter, Knell Family, Peter and Patricia Gruber Award, the Hamburger Family, Karl-Johan Persson and Miel de Botton

AUTHOR CONTRIBUTIONS

A.N., S.K., and Y.S. designed the study and wrote the paper. A.N., S.K., and C.B. analyzed the data and performed and supervised the experiments. A.N., S.K., M.A., and P.G. performed the HLA peptidomics experiments. A.N., P.G., G.Y., and A.P. grew the cell lines for HLA peptidomics. Y.L. helped with the HLA peptidomics and mass spectrometry analyses. G.B. performed HLA typing. Y.Y.-R., L.E., N.E., E.B.-H., and N.P. performed SARS-CoV-2 infection experiments. A.G., S.L.C.K., and P.K. performed T cell reactivity assays. J.P., A.N., S.K., and Y.S. designed the study and wrote the paper. A.N., S.K., and Y.S. designed the study and wrote the paper.

DECLARATION OF INTERESTS

Patent applications have been filed on SARS-CoV-2 peptides (Y.S., A.N., and S.K.). The remaining authors declare no competing interests.

Received: November 3, 2020
Revised: January 17, 2021
Accepted: June 3, 2021
Published: June 29, 2021

REFERENCES


Figure 5. Presented SARS-CoV-2 peptides are immunogenic

CD8+ T cell recognition was assessed for the identified HLA-I peptides by using fluorescent pHLA multimers. Flow cytometry plots of detected SARS-CoV-2-specific CD8+ T cell responses in COVID-19 patients or healthy non-exposed controls. The magnitude of the response is defined as the percentage of double-positive pHLA+ cells of total CD8+ cells.


Nielsen, M., and Andreatta, M. (2016). NetMHCpan-3.0; improved prediction of binding to MHC class II molecules integrating information from multiple receptors and peptide length datasets. Genome Med. 9, 11.


### Key Resources Table

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse monoclonal anti-pan HLA-I</td>
<td>Purified from HB95 hybridoma cells</td>
<td>N/A</td>
</tr>
<tr>
<td>(clone W6/32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse monoclonal anti-pan HLA-II</td>
<td>Purified from HB145 hybridoma cells</td>
<td>N/A</td>
</tr>
<tr>
<td>(clone IVA12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse monoclonal anti-human ACE2</td>
<td>Santa Cruz</td>
<td>Cat# sc-390851: RRID:AB_2861379</td>
</tr>
<tr>
<td>(clone E-11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse monoclonal anti-human TMPRSS2</td>
<td>Santa Cruz</td>
<td>Cat# sc-515727</td>
</tr>
<tr>
<td>(clone H4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alexa Fluor® 488 Goat polyclonal anti-</td>
<td>Jackson ImmunoResearch</td>
<td>Cat#115-545-146: RRID:AB_2307324</td>
</tr>
<tr>
<td>mouse IgG</td>
<td></td>
<td></td>
</tr>
<tr>
<td>APC-Streptavidin</td>
<td>Invitrogen</td>
<td>Cat#S868</td>
</tr>
<tr>
<td>BUV496 Mouse monoclonal anti-human</td>
<td>BD</td>
<td>Cat#612944: RRID:AB_2870224</td>
</tr>
<tr>
<td>CD16 (clone 3G8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BU563-Streptavidin</td>
<td>BD</td>
<td>Cat#567655</td>
</tr>
<tr>
<td>BU615-Streptavidin</td>
<td>BD</td>
<td>Cat#13013</td>
</tr>
<tr>
<td>BU661 Mouse monoclonal anti-human</td>
<td>BD</td>
<td>Cat#750536: RRID:AB_2874685</td>
</tr>
<tr>
<td>CD19 (clone SJ25C1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BU805 Mouse monoclonal anti-human</td>
<td>BD</td>
<td>Cat#612889: RRID:AB_2833078</td>
</tr>
<tr>
<td>CD8 (clone SK1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BV650-Streptavidin</td>
<td>BD</td>
<td>Cat#563855</td>
</tr>
<tr>
<td>FITC Mouse monoclonal anti-human CD14</td>
<td>BD</td>
<td>Cat#345784: RRID:AB_2868810</td>
</tr>
<tr>
<td>(clone M9P9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LIVE/DEAD Fixable Near-IR Dead Cell Stain Kit</td>
<td>Invitrogen</td>
<td>Cat#L10119</td>
</tr>
<tr>
<td><strong>Bacterial and virus strains</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pCDH-CMV-MCS-EF1α-Neo vector</td>
<td>System Biosciences</td>
<td>Cat# CD514B-1</td>
</tr>
<tr>
<td>pLVX-EF1alpha-ires-Puro</td>
<td>Gordon et al., 2020</td>
<td>N/A</td>
</tr>
<tr>
<td>SARS-CoV-2 virus- EPI_ISR_406862</td>
<td>Bundeswehr Institute of Microbiology, Munich, Germany</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Chemicals, peptides, and recombinant proteins</strong></td>
<td>GenScript</td>
<td>N/A</td>
</tr>
<tr>
<td>Custom made synthetic peptides</td>
<td>Sigma</td>
<td>Cat#P8340</td>
</tr>
<tr>
<td>Protease Inhibitors Cocktail</td>
<td>Sigma</td>
<td>Cat#D6750</td>
</tr>
<tr>
<td>sodium deoxycholate</td>
<td>Sigma</td>
<td>Cat#6125</td>
</tr>
<tr>
<td>iodoacetamide</td>
<td>Sigma</td>
<td>Cat#V4231</td>
</tr>
<tr>
<td>EDTA</td>
<td>Promega</td>
<td>Cat#78830</td>
</tr>
<tr>
<td>PMSF</td>
<td>Sigma</td>
<td>Cat#O8001</td>
</tr>
<tr>
<td>octyl-β-D glucopyranoside</td>
<td>Sigma</td>
<td>Cat#L00210</td>
</tr>
<tr>
<td>Protein-A Resin</td>
<td>AβS</td>
<td></td>
</tr>
<tr>
<td><strong>Deposited data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw MS files</td>
<td>ProteomeXchange via PRIDE</td>
<td>PXD023614</td>
</tr>
<tr>
<td><strong>Experimental models: cell lines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IHW01070</td>
<td>IHWG Cell and DNA Bank</td>
<td>N/A</td>
</tr>
<tr>
<td>IHW01161</td>
<td>IHWG Cell and DNA Bank</td>
<td>N/A</td>
</tr>
<tr>
<td>LCL 721.221</td>
<td>ATCC</td>
<td>CRL-1855</td>
</tr>
<tr>
<td>Calu-3</td>
<td>ATCC</td>
<td>HTB-55</td>
</tr>
</tbody>
</table>

(Continued on next page)
### RESOURCE AVAILABILITY

#### Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Prof. Yardena Samuels [yardena.samuels@weizmann.ac.il](mailto:yardena.samuels@weizmann.ac.il).

#### Materials availability

SARS-CoV-2 envelope, nuclocapsid, nonstructural protein-6 and membrane protein, pLVX-EF1alpha-IRES-Puro plasmids were a kind gift from Prof. Nevan J. Krogan ([Gordon et al., 2020](https://doi.org/10.1016/j.celrep.2020.108355)).

### Table: REAGENT or RESOURCE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calu-6</td>
<td>ATCC</td>
<td>HTB-56</td>
</tr>
<tr>
<td>HEK293T/AE2</td>
<td>GenScript</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Oligonucleotides</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primers RT E: F- TCAGAAGAAAACC GGGACACT</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Primers RT E: R- TGCCAGAAACAA GAGCAG</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Primers RT N: F- CGAGGACAGGG TGTACCAAT</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Primers RT N: R- ACCATCTCCACC TCTGATGC</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Primers RT nsp6: F- CGACCAGGGCT ATTTCCATGT</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Primers RT nsp6: R- CCCTCTCGCC AAAAACATAA</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Primers RT M: F- TATTCCTTTTGCC TCTGTGGG</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Primers RT M: R- GCCGCCAGGTA TCCAGTTTA</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Software and algorithms</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FlowJo</td>
<td>FlowJo, LLC</td>
<td>N/A</td>
</tr>
<tr>
<td>MaxQuant version 1.5.0.25</td>
<td>Cox and Mann 2008</td>
<td>N/A</td>
</tr>
<tr>
<td>NetMHCpan version 4.1</td>
<td>Hoof et al., 2009</td>
<td><a href="http://www.cbs.dtu.dk/services/NetMHCpan/">http://www.cbs.dtu.dk/services/NetMHCpan/</a></td>
</tr>
<tr>
<td>NetMHCIIpan version 4</td>
<td>Jensen et al., 2018</td>
<td><a href="http://www.cbs.dtu.dk/services/NetMHCIIpan/">http://www.cbs.dtu.dk/services/NetMHCIIpan/</a></td>
</tr>
<tr>
<td>MixMHCpred</td>
<td>Bassani-Sternberg et al., 2017</td>
<td><a href="https://github.com/GfellerLab/MixMHCpred">https://github.com/GfellerLab/MixMHCpred</a></td>
</tr>
<tr>
<td>HL AthenA</td>
<td>Abelin et al., 2017</td>
<td><a href="http://hiathena.tools/">http://hiathena.tools/</a></td>
</tr>
<tr>
<td>NeonMHC2</td>
<td>Abelin et al., 2019</td>
<td><a href="https://neonmhc2.org/neonmhc2/neonmhc2_main/">https://neonmhc2.org/neonmhc2/neonmhc2_main/</a></td>
</tr>
<tr>
<td>R package specL v1.6.2</td>
<td>Krokchin et al., 2004</td>
<td><a href="https://www.rdocumentation.org/packages/specL/versions/1.6.2">https://www.rdocumentation.org/packages/specL/versions/1.6.2</a></td>
</tr>
<tr>
<td>GibbsCluster 2.0 server</td>
<td>Andreattta et al., 2017</td>
<td><a href="https://www.cbs.dtu.dk/services/GibbsCluster">https://www.cbs.dtu.dk/services/GibbsCluster</a></td>
</tr>
<tr>
<td>Seq2Logo 2.0</td>
<td>Thomsen and Nielsen. 2012</td>
<td><a href="http://www.cbs.dtu.dk/biotools/Seq2Logo">http://www.cbs.dtu.dk/biotools/Seq2Logo</a></td>
</tr>
<tr>
<td>Immune epitope database (JEDB)</td>
<td>Vita et al., 2019</td>
<td><a href="https://www.iedb.org">https://www.iedb.org</a></td>
</tr>
<tr>
<td>Panther classification system version 15.0</td>
<td>Mi et al., 2019</td>
<td><a href="http://pantherdb.org/">http://pantherdb.org/</a></td>
</tr>
</tbody>
</table>
Data and code availability
All raw MS files as well as human and viral proteomes used for analyzing the data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD025499. All the datasets are related to Figures 2, 3, 4, and S1. Analyzed HLA peptidomics data and additional figures and tables were deposited in Database: https://github.com/ShellyKalaora/Identification-of-presented-SARS-CoV-2-HLA-I-and-HLA-II-peptides-using-HLA-peptidomics.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

SARS-CoV-2 virus
SARS-CoV-2 virus (BetaCoV/Germany/BavPat1/2020 EPI_ISL_406862, GISAID Acc. No. EPI_ISL_406862) was kindly provided by Bundeswehr Institute of Microbiology, Munich, Germany. The virus was propagated (total 4 passages) and titered on Vero E6 cells (Finkel et al., 2020a; Yahalom-Ronen et al., 2020).
Handling and working with SARS-CoV-2 virus was conducted in a BSL3 facility in accordance with the biosafety guidelines of the Israel Institute for Biological Research. The Institutional Biosafety Committee of Weizmann Institute approved the protocol used in these studies.

Human samples
Peripheral blood samples from COVID-19 patients were collected in accordance with the Declaration of Helsinki after approval by the institutional review boards (Ethical Committee of Area Vasta Emilia Romagna, protocol number 177/2020, March 10th, 2020, and subsequent amendments). Each participant signed informed consent. All COVID-19 patients were tested positive for SARS-CoV-2 using reverse transcriptase chain reaction (RT-PCR) from an upper respiratory tract (nose/throat) swab test in accredited laboratories. Peripheral blood samples were obtained during hospitalization from COVID-19 patients with critical and severe disease, and three months after confirmed infection from asymptomatic COVID-19 patients.

METHOD DETAILS

Selection of most frequent HLA-I alleles in the world population
HLA-I alleles frequency in the world population was downloaded for each available country from the Allele frequency net database (http://www.allelefrequencies.net/). All alleles with frequency above 0.05 were kept for further analysis, and the average frequency for each allele was calculated. The top 8 HLA-A (A*01:01, A*02:01, A*03:01, A*11:01, A*24:02, A*68:01, A*23:01 and A*33:03), 6 HLA-B (B*07:02, B*08:01, B*18:01, B*35:01, B*40:01 and B*51:01) and 6 HLA-C (C*01:02, C*03:04, C*04:01, C*06:02, C*07:01 and C*07:02) alleles that have the highest average frequency and found in the highest number of populations were selected and used for further analyses. Together these alleles cover at least one of the HLA-A/B/C top allele of each population, and in most cases more than one of the top alleles. NetMHCpan (Hoof et al., 2009; Jurtz et al., 2017; Nielsen and Andreatta, 2016) was used to predict all HLA-I peptides from the SARS-CoV-2 proteome for all selected frequent alleles. (This analysis is related to Figure S1).

Cells lines
EBV-transformed B cells (IHW01070 and IHW01161) were purchased from the IHWG Cell and DNA Bank. LCL 721.221 HLA-I null cells and Calu-6 cells were purchased from the ATCC. HEK293T stably expressing ACE2 were a kind gift from Prof. Ron Diskin, Calu-3 cells were a kind gift from Dr. Noam Stern-Ginossar. All cell lines were tested regularly and were found negative for mycoplasma contamination (EZ-PCR Mycoplasma Kit, Biological Industries). All B cell lines and Calu-6 cells were maintained in RPMI 1640 containing 10% fetal bovine serum (FBS) and 1% penicillin-streptomycin, glutamine and sodium pyruvate. HEK293T and Calu-3 were maintained in DMEM containing 10% FBS and 1% penicillin-streptomycin and glutamine.

HLA typing
Genomic DNA for 721.221 cell line, Calu-6 cells and peripheral blood samples from COVID-19 patients was extracted from 2*10^6 cells. DNA samples were typed for six loci: HLA-A, -B, -C, -DPB1, -DQB1 and DRB1, using the MX6-1 NGS typing kit (GenDx). HLA typing for IHW01070 and IHW01161 cells is from IHWG Cell and DNA Bank database. HLA typing information is found in Figure S1D. (This data is related to Figures 3, 4, S1, and S2).

Pooled stable expression of HLA-I alleles in 721.221 cells
DNA sequences coding for HLA-I alleles were taken from the IPD-IMGT/HLA database (https://www.ebi.ac.uk/ipd/imgt/hla/allele.html) and purchased as synthetic dsDNA from Twist bioscience. Coding sequence was cloned into pCDH-CMV-MCS-EF1α-Neo vector (SBI, #CD514B-1) and lentiviral particles were produced by co-transfection with envelope and packaging plasmids (PMD2.G and psPAX2) into HEK293T cells using Lipofectamine 2000 (Invitrogen). At 48 hours post transfection, the virus containing media was harvested, filtered, aliquoted and stored at −80°C. Human B-LCL 721.221 (HLA-I null) were infected, and after 72 hours were selected with neomycin (G418). (This experiment is related to Figures 3, 4, and S1).
Pooled stable expression SARS-CoV-2 protein expressing B cells and Calu-6 cells

To produce the lentivirus of SARS-CoV-2 envelope, nuclocapsid, nonstructural protein-6 and membrane protein, pLVX-EF1alpha-IRES-Puro plasmids (Gordon et al., 2020) were co-transfected with pCMV-VSV-G and psPAX2 helper plasmids using Lipofectamine 2000 (Life Technologies) into HEK293T-cells. The cells were seeded at 2.5 x 10^6 per T75 flask. Virus-containing medium was collected 72 hours after transfection, filtered, aliquotted and stored at -80°C. IHV1161, IHV1070 and 721.221 mono-allelic B cells or Calu-6 cells were infected with the virus for 48 hours and then selected with Puromycin. The expression of the viral genes was confirmed by quantitative PCR assay. We observed variable expression levels of the genes in some of the cell lines, which might limit peptide identification in these cells. (This experiment is related to Figures 3, 4, and S1).

Quantitative PCR assay

Total RNA was extracted from cells by using RNeasy Mini Kit (Cat: 74107, QIAGEN), which was then converted into cDNA using iScript reverse Transcription Supermix (Bio-Rad), according to the manufacturer’s instruction. Real-time PCR analysis was performed in triplicate using Fast SYBR Green Master Mix (cat: 4385612, Appliedbiosystems). The reaction condition was 95°C for 5 min, followed by 40 cycles of denaturation at 95°C for 15 s and annealing/elongation step at 60°C for 30 s. The relative expression was analyzed by the 2^−ΔΔCT method. Envelope: F- TCAGAAGAAACCUGGACACT, R- TGCCAGAAGAAACGACACAG
Nucleocapsid: F- CGAGGACAGGGTGTAACAAAT, R- ACCATCTCCACCTCTGATGC, nsp6: F- CGACCAGCTATTCCATGT
R- CCCCCTCGCAAATACATAAA, Membrane: F- TATTCCTTTGTGCTTGTGG, R- GCGCCAGTTATCCAGTTTA. (This experiment is related to Figures 3, 4, and S1).

Production and purification of membrane HLA molecules

Cell pellets were homogenized by pipetting on ice with a lysis buffer containing 0.25% sodium deoxycholate, 0.2 mM iodoacetamide, 1 mM EDTA, 1:200 Protease Inhibitor Cocktail (Sigma-Aldrich, P8340), 1 mM PMSF and 1% octyl-β-D glucopyranoside in PBS. Samples were then incubated in rotation at 4°C for 1 hour. The lysates were cleared by centrifugation at 48,000 g for 60 minutes at 4°C and then passed through a pre-clearing column containing Protein-A Sepharose beads. HLA-I molecules were immunoaffinity purified from cleared lysate with the pan-HLA-I antibody (W6/32 antibody purified from HB95 hybridoma cells) covalently bound to Protein-A Sepharose beads. HLA-II molecules were then transfused with the flow-through to similar affinity columns containing a pan-HLA-II antibody (purified from HB-145 hybridoma cells). Affinity columns were washed first with 400 mM NaCl, 20 mM Tris–HCl and then with 20 mM Tris–HCl pH 8.0. The HLA-peptide complexes were then eluted with 1% trifluoracetic acid followed by separation of the peptides from the proteins by binding the eluted fraction to Sep-Pak (Waters). Elution of the peptides was done with 28% acetonitrile in 0.1% trifluoracetic acid for HLA-I and 32% acetonitrile in 0.1% trifluoracetic acid for HLA-II. (This experiment is related to Figures 2, 3, 4, S1, and S2).

Mass-spectrometry analysis of eluted HLA peptides

UPLC/MS grade solvents were used for all chromatographic steps. Each sample was solubilized in 12 μL 97:3 water: acetonitrile with 0.1% formic acid. Samples were loaded using split-less nano-Ultra Performance Liquid Chromatography (10 kpsi nanoAcquity; Waters, Milford, MA, USA). The mobile phase was: A) H2O + 0.1% formic acid and B) acetonitrile + 0.1% formic acid. Desalting of the samples was performed online using a reversed-phase Symmetry C18 trapping column (180 μm internal diameter, 20 mm length, 5 μm particle size, Waters). The peptides were then separated using a T3 HSS nano-column (75 μm internal diameter, 250 mm length, 1.8 μm particle size; Waters) at 0.35 μL/minute. Peptides were eluted from the column into the mass spectrometer using the following gradient: 5% to 28% B in 120 minutes, 28% to 35% B in 15 minutes, 35% to 95% in 15 minutes, maintained at 95% for 10 minutes and then back to initial conditions.

The nanoUPLC was coupled online through a nanoESI emitter (10 μm tip; New Objective, Woburn, MA, USA) to a quadrupole orbitrap mass spectrometer (Q Exactive Plus, Thermo Scientific) using a FlexIon nanospray apparatus (Proxeon).

For the discovery experiments, data was acquired in data dependent acquisition (DDA) mode, using a Top20 method. MS1 resolution was set to 17,500, quadrupole isolation 1.8 m/z, AGC of 1e5, dynamic exclusion of 20 s and maximum injection time of 150 msec. Charge exclusion was set to ‘unassigned’, 4-8, greater than 8 and an exclusion list of singly charged, background ions.

Identification of SARS-CoV-2-derived HLA peptides

SARS-CoV-2 proteome FASTA file of the infection experiment was generated according to the genome of the virus used for the infections (BetaCoV/Germany/BavPat1/2020 EPI_ISR_406862, GISAID Acc. No. EPI_ISR_406862), as well as the 23 novel unannotated virus ORFs whose translation is supported by Ribo-seq (Finkel et al., 2020a). For the overexpressed genes, the protein sequences used derived from the SARS-CoV-2 reference genome (Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, NCBI Reference Sequence: NC_045512.2). Human proteome was downloaded from UniProt (https://www.uniprot.org/). MS data was analyzed using MaxQuant (Cox and Mann, 2008) version 1.5.0.25 with the human and viral proteome in the same run. Enzyme specificity was set as “unspecific” and peptides’ FDR was set to 0.05. The “match between runs” option was disabled to avoid the matching of identifications across the samples.
For the identification of SARS-CoV-2 derived peptides, as Leucine and Isoleucine residues are generally considered indistinguishable by MS, because their molecular masses are the same, we removed from our list all peptides in which replacing their I/L position (by either I/L) resulted in peptides that can be derived from a human origin. We also removed peptides that have the same sequence (or with a change of I/L) as possible non-coding regions (https://www.gencodegenes.org/human/release_19.html) as well as possible pseudogenes (http://www.pseudogene.org/Human/Human90.txt). NetMHCpan (Hoof et al., 2009; Jurtz et al., 2017; Nielsen and Andreatta, 2016) version 4.1 (http://www.cbs.dtu.dk/services/NetMHCpan/) and NetMHCIclip (Jensen et al., 2018) version 4.0 (http://www.cbs.dtu.dk/services/NetMHCIclip/) were used to check if the peptides can bind the patient’s HLA alleles. Peptides with % rank = < 2 or = < 10 were kept for HLA-I and HLA-II, respectively. We predicted the binding of SARS-CoV-2 derived peptides also by MixMHCpred (peptide length: 8-14 for HLA-I and > = 9 aa for HLA-II, peptides with % rank = < 2 were determined as binders) (Bassani-Sternberg et al., 2017; Gfeller et al., 2018), HL Athena (HLA-I peptides, peptide length: 8-11, binders were determined according to columns ‘best.MSi_allele’ and ‘assign.MSi_allele’, % rank = < 10) (Abelin et al., 2017; Sarkizova et al., 2020) and NeonMHC2 (HLA-II peptides, peptide length: > = 9, % rank ≤ 10) (Abelin et al., 2019).

We validated the identification of randomly selected peptides by comparing their MS/MS spectra fragmentation to that of synthetic peptides (see STAR Methods: “Validation using synthetic peptides”). We also clustered human and SARS-CoV-2 derived peptides using Gibbs clustering, to see if they clustered according to the expected HLA binding motifs of the patient (see STAR Methods: “Gibbs clustering”). (This experiment is related to Figures 2, 3, 4, S1, and S2).

Hydropophicity index calculation

Sequence specific hydropophicity index was calculated using the ssrc function from the R package specL v1.6.2 (Krokhin et al., 2004) with default parameters. Observed retention times (RT) were obtained from MaxQuant output file “smsms.txt.” Peptides’ RTs were plotted against the calculated hydropophicity index. We regressed the measured RTs against the calculated hydropophicity index using the lm function from the R package stats v3.6.2, in order to calculate the standard errors of the hydropophicity index. The residual absolute errors of the lm-regression were plotted using a boxplot with a median and hinges (25% and 75%). The outliers were determined as residual absolute errors greater than 75th percentile + 1.5 * interquartile range (IQR). Figures are available in https://github.com/ShellyKalaora/Identification-of-presented-SARS-CoV-2-HLA-I-and-HLA-II-peptides-using-HLA-peptidomics.

Validation using synthetic peptides

Synthetic peptides (GenScript, 50 fmol/μl) were used to validate the peptides’ fragmentation. Peptides were analyzed in the same MS and conditions as the eluted peptide samples from cell lines. The MSnbase R package (Gatto and Lilley, 2012) was used to calculate the correlation between the matched y and b ions of the synthetic peptides and the endogenous peptides. Pearson correlation and dot product score are indicated for each comparison in the figure. SpectrumSimilarity function from the R package OrgMassSpecR was used to plot a head to tail figure of the endogenous and synthetic peptide fragmentation. Figures are available in https://github.com/ShellyKalaora/Identification-of-presented-SARS-CoV-2-HLA-I-and-HLA-II-peptides-using-HLA-peptidomics.

Gibbs clustering

Each set of peptides was clustered using the GibbsCluster 2.0 server (Andreatta et al., 2013, 2017; https://www.cbs.dtu.dk/services/GibbsCluster), with the “MHC class I ligands of length 8-13” parameters; the number of clusters was set to six and the trash cluster option was enabled. Since the number of peptides per allele is different, for alleles with a higher number of peptides (as the HLA-A alleles), the unbiased clustering sometimes resulted in more than one cluster for these alleles. In these cases, we selected the number of clusters in which we receive only one cluster per HLA allele. The variation in the number of peptides per allele also resulted in clusters with mixed motifs that were similar. In these cases, we assigned the cluster to the allele that had the highest representation in the cluster and added a note as to which other alleles are mixed within. For each cluster, we indicated the number of human- and SARS-CoV-2-derived peptides clustered to this allele. All motifs were generated by Seq2Logo 2.0 (Thomsen and Nielsen, 2012; http://www.cbs.dtu.dk/biotools/Seq2Logo) with the default settings. In order to classify the clustered peptides into HLA alleles, we first identified for each allele its motif. This entailed retrieving all HLA-I epitopes registered under this allele from the Immune epitope database (IEDB; https://www.iedb.org; Vita et al., 2019). All peptides annotated as positive in “MHC ligand assays” to the specific HLA-I allele and were 8-13 amino acids in length. The GibbsCluster 2.0 server was used to align the peptides using the “MHC class I ligands of length 8-13” parameters; the number of clusters was set to one and the trash cluster option was disabled. HLA alleles to which there were no peptides in the IEDB to create a representative logo for their motif were searched in HL Athena.tools (Sarkizova et al., 2020). Figures are available in https://github.com/ShellyKalaora/Identification-of-presented-SARS-CoV-2-HLA-I-and-HLA-II-peptides-using-HLA-peptidomics.

Flow cytometry

Cells were stained with mouse monoclonal IgG1 anti-ACE2 antibody (cat: sc-390851, clone E-11, lot: D2420, Santa Cruz) or anti-TMPRSS2 (cat: sc-515727, clone:H-4, lot:D2420), followed by Alexa Fluor® 488 AffiniPure Goat Anti-Mouse IgG (cat: 115-545-146, lot: 138610). Flow cytometer (BD Biosciences) was used and the data was analyzed using FlowJo software (FlowJo, LLC).
SARS-CoV-2 viral infection

2x10^8 cells were centrifuged at 300 g for 5 minutes and washed once with RPMI without fetal bovine serum (FBS). IHW01070, HEK293T cell pellets were infected with SARS-CoV-2 virus at a multiplicity of infection (MOI) of 0.05, 0.7 and 3, respectively, in RPMI medium supplemented with 2% fetal bovine serum (FBS), MEM non-essential amino acids, 2mM L-Glutamine, 100Units/ml Penicillin, 1% non-essential amino acid, 1% Na-pyruvate and 20 µg per ml TPCK trypsin (Thermo scientific) at a final volume of 2 mL for 1 hour at 37°C with gentle agitation every 15 minutes. After 1 hour of infection, additional 20 mL of similar infection medium without TPCK trypsin were added and the infected cells and plated in two T-75 flasks for 24 hours in CO2 humidified incubator at 37°C, 5% CO2. Following the 24 hours of infection, cells were centrifuged (300 g, 5 minutes,) washed once with PBS and the cell pellet was stored at −70°C. Cell pellets were thawed, suspended in lysis buffer for 1 hour and stored on ice for 1 hour. The Calu-3 cells were infected as previously described (Finkel, 2021) (This experiment is related to Figures 2, 3, 4, and S2).

SARS-CoV-2 cell infection efficiency

Immunofluorescence staining of IHW1070, HEK293T cell lines was preformed using SARS CoV-2 specific antibodies (Finkel et al., 2020a; Yahalom-Ronen et al., 2020). Briefly, Cells were infected as described above and 24 hours later 1x10^8 cells were harvested by centrifugation (300 g, 5 minutes), washed once with PBS and fixed with 3% paraformaldehyde (PFA) in PBS for 20 minutes. Cells were permeabilized with 0.5% Triton X-100 for 2 minutes, blocked with PBS containing 2% FBS and stained with hyperimmune rabbit serum from intravenous (i.v.) SARS-CoV-2 infected rabbits, for 30 minutes, washed with PBS, and incubated with Alexa Fluor 488-conjugated secondary antibody. Nuclei were visualized by staining with 5µg/ml of 4,6-Diamidino-2-Phenylindole (DAPI). The Immunofluorescence staining of Calu-3 cells was previously described (Finkel, 2021). Representative confocal images were acquired with a Zeiss LSM 800 confocal microscope using a 20x objective and processed with ImageJ (NIH). Images were taken of three different fields and used for infection rate calculation. (This experiment is related to Figure S2).

Proteomic analysis of infected cells

The flow through of the HLA peptidomics lysates was used for proteomic analysis. The proteins in a mixture with 8 M urea and 100 mM ammonium bicarbonate were reduced and digested in 2 M urea, 25 mM ammonium bicarbonate with modified trypsin or chymotrypsin (Promega) at a 1:50 enzyme-to-substrate ratio. The resultant peptides were desalted using C18 tips (Homemade stage tips) and subjected to LC-MS-MS analysis. The peptides were resolved by reverse-phase chromatography on 0.075 x 180-mm fused silica capillaries (J&W) packed with Reprosil reversed phase material (Dr. Maisch GmbH). The peptides were eluted with a linear 30-min gradient of 5%–95% acetonitrile with 0.1% formic acid in water, 15-min gradient of 35%–95% acetonitrile with 0.1% formic acid in water, and 15 min at 95% acetonitrile with 0.1% formic acid in water at a flow rate of 0.15 µL/min. Mass spectrometry was performed with a Q Exactive plus mass spectrometer (Thermo) in a positive mode using repetitively full MS scan, followed by high-energy collision-induced dissociation (HCD) of the 10 most dominant ions selected from the first MS scan. The mass spectrometry data were analyzed using MaxQuant version 1.5.2.8 against a human UniProt database with a mass tolerance of 10 ppm for the precursor masses and 0.05 amu for the fragment ions. Oxidation on Met was accepted as a variable modification, and carbamidomethyl data were analyzed using MaxQuant version 1.5.2.8 against a human UniProt database with a mass tolerance of 10 ppm for the precursor masses and 0.05 amu for the fragment ions. Oxidation on Met was accepted as a variable modification, and carbamidomethyl (This analysis is related to Figure 2).

Analysis of differentially expressed proteins and presented peptides

HLA peptides and proteins identified through MaxQuant were first filtered to remove reverse sequences and known contaminants. Only HLA peptides that were predicted to bind the cell’s HLA alleles were used for this analysis. Graphics and statistical analysis was done using the Perseus computational platform version 1.6.6.0 (Tyanova et al., 2016). HLA peptides intensities or proteins LFQ intensity were Log2 transformed and peptides/proteins with at least 2 values in one group (infected or non-infected) were kept. Missing intensity values were imputed by drawing random numbers from a Gaussian distribution with a standard deviation of 30% in comparison to the standard deviation of the measured peptide abundances (width 0.3 and downshift 1.8). Volcano plots which show differentially presented peptides or expressed proteins after SARS-CoV-2 infection were plotted, and the x axis represented the Log2 fold changes of the peptide intensities, and the y axis represent the significance levels calculated by two-sided unpaired t test with a FDR of 0.05 and S0 of 1 for HLA peptides or S0 of 0.1 for proteins. (This analysis is related to Figure 2).

Pathway analysis

HLA peptides that were found to be more presented were query to identify enrichment of pathways. The list was uploaded to the Panther classification system version 15.0 website (Mi et al., 2019) and statistical over-representation test was preformed against the Panther pathways annotation set with human genes list as a reference. A Fisher exact test was used to find over represented pathways with FDR < 0.05. (This analysis is related to Figure 2).
Similarity of SARS-CoV-2 peptides to other CoV species

The peptides identified in the study were searched for similarity to the proteome of the following CoV family species (Decaro and Lorusso, 2020): HCoV-NL63, HCoV-229E, HCoV-OC43, HCoV-HKU1, SARS-CoV (SARS-CoV-1) and MERS-CoV. We searched for peptides that have identical sequence or peptides with one amino acid substitution. (This analysis is related to Figure 4).

Finding SNPs in the SARS-CoV-2 genome

All the SARS-CoV-2 genomes from the GISAID database (https://www.gisaid.org/) were downloaded with the following requirements: complete genome, high coverage and human as host. In total 211,763 sequences were used for the analysis. All sequences were aligned to the reference sequence using the EMBOSS stretcher. Only SNPs resulted in amino acid change were counted, and their percentage was counted from total number of sequences. (This analysis is related to Figure 2).

Similarity of SARS-CoV-2 peptides to human proteome

The peptides identified in the study as well as all predicted HLA-I peptides matching the most frequent alleles were searched for similarity to the human proteome. We searched for peptides that have identical sequence or peptides with one amino acid substitution. Relevant table is available in https://github.com/ShellyKalaora/Identification-of-presented-SARS-CoV-2-HLA-I-and-HLA-II-peptides-using-HLA-peptidomics.

Generation of fluorescent pHLA multimers

MHC complexes were loaded with the peptides of interest via UV-induced peptide exchange, as described previously (Hadrup et al., 2009; Rodenko et al., 2006). Different fluorescent streptavidin (SA) conjugates were added to 10 μL of pHLA monomer (100 μg/ml): 0.6 μL of SA-APC (Invitrogen, S868), 2 μL of SA-BV650 (BD, 563855), 1 μL of SA-BUV615 (BD, 613013), 1.5 μL of SA-BUV563 (BD, 565765). For each pHLA monomer, conjugation was performed with two of the fluorescent SA conjugates. Next, milk (1% w/v, Sigma) was added to block unspecific peptide binding residues. After 30 minutes of incubating on ice, D-biotin (26.3 mM, Sigma) in PBS and NaN3 (0.02% w/v) was added to block residual biotin binding sites. The fluorescent pHLA multimers were left overnight at 4°C before using them. (This experiment is related to Figures 5 and S3).

Surface staining with pHLA multimers and antibodies

PBMCs were thawed, washed and re-suspended in 1 mL complete RPMI (cRPMI; RPMI 1640 supplemented with 10% Human Serum and 1% Penicillin-Streptomycin) and Benzonase nuclease (Merck-Millipore, 2500 IU/ml) and incubated at 37°C for 30 minutes. The cells were washed and stained with the following amounts of fluorescently labeled pHLA multimers: 2 μL of SA-APC-pHLA and 1 μL SA-BV650-pHLA, SA-BUV615-pHLA, SA-BUV563-pHLA. The cells were stained in 100 μL of Brilliant Staining Buffer Plus (BD, 563794) according to manufacturer’s protocol. After 15 minutes of incubating at 37°C, the cells were stained with 2 μL of aCD4-BB700 (BD, 566392), 0.5 μL of aCD14-FITC (BD, 345784), 1 μL of aCD16-BUV496 (BD, 612944), 0.5 μL aCD19-BUV661 (BD, 750536) and 0.5 μL of LIVE/DEAD Fixable IR Dead Cell Stain Kit (Invitrogen, L10119) and incubated on ice for 20 minutes. Samples were analyzed on the BD FACSymphony A5. The following gating strategy was applied to identify CD8+ T cells: (i) selection of live (IR dye low-dim) single-cell lymphocytes (forward scatter (FSC)-W/H low, side scatter (SSC)-W/H low, FSC/SSC-A), (ii) selection of aCD4, aCD14, aCD16, aCD19 negative and aCD8+ cells. Antigen-specific CD8+ T cell responses that were positive for two and none of the other pHLA multimer channels were identified using Boolean gating. Cut-off values defining true positive responses were ≥ 0.005% of total CD8+ T cells, ≥ 5 events. A minimum of 10,000 CD8+ T cells was acquired per sample. Data was analyzed using FlowJo 10.6.2. (This experiment is related to Figures 5 and S3).

Flow cytometer setting for pHLA multimers

The following 21-color instrument settings were used on the BD FACSymphony A5: blue laser (488 nm at 200 mW): FITC, 530/30BP; 530/30BP, 505LP; BB630, 600LP, 610/20BP; BB700, 710/50BP, 685LP; BB790, 750LP, 780/60BP. Red laser (637 nm at 140 mW): APC, 670/30BP, APC-R700, 690LP, 630/45BP, IRDye, 750LP, 780/60BP, Violet laser (405 nm at 100 mW): BV421, 420LP, 431/28BP; BV480, 455LP, 470/20BP; BV650, 565LP, 605/40BP; BV650, 635LP, 661/11BP; BV711, 711/85, 685; BV750, 735LP, 750/30BP, UV laser 355 nm at 75 mW: BV395, 379/28BP, BV450, 525LP, 580/20BP; BV563, 550LP, 580/20BP; BV615, 600LP, 615/20BP; BV661, 630LP, 670/25BP; BV780, 770LP, 819/44BP. Yellow-green laser (561 nm at 150 mW): PE, 586/15BP. Appropriate compensation controls were included in each analysis.

QUANTIFICATION AND STATISTICAL ANALYSIS

In volcano plots showing differentially presented peptides or expressed proteins after SARS-CoV-2 infection the significance levels were calculated by two-sided unpaired t test with a FDR of 0.05 and S0 of 1 for HLA peptides or S0 of 0.1 for proteins.