



OPEN ACCESS

EDITED BY
Markus Cornberg,
Hannover Medical School, Germany

REVIEWED BY
Dario Gherzi,
University of Nebraska Omaha,
United States
Luzia Maria De-Oliveira-Pinto,
Oswaldo Cruz Foundation (Fiocruz),
Brazil

*CORRESPONDENCE
Gustavo Fioravanti Vieira
gustavo.vieira@unilasalle.edu.br

[†]These authors have contributed
equally to this work and share
first authorship

SPECIALTY SECTION
This article was submitted to
T Cell Biology,
a section of the journal
Frontiers in Immunology

RECEIVED 28 April 2022

ACCEPTED 04 July 2022

PUBLISHED 28 July 2022

CITATION
Antonio EC, Meireles MR,
Bragatte MAdS and Vieira GF (2022)
Viral immunogenic footprints
conferring T cell cross-protection to
SARS-CoV-2 and its variants.
Front. Immunol. 13:931372.
doi: 10.3389/fimmu.2022.931372

COPYRIGHT
© 2022 Antonio, Meireles, Bragatte and
Vieira. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](#). The use, distribution or
reproduction in other forums is
permitted, provided the original author
(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Viral immunogenic footprints conferring T cell cross- protection to SARS-CoV-2 and its variants

Eduardo Cheuiche Antonio^{1†}, Mariana Rost Meireles^{1†}, Marcelo
Alves de Souza Bragatte¹ and Gustavo Fioravanti Vieira^{1,2*}

¹Post Graduation Program in Genetics and Molecular Biology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, ²Post Graduation Program in Health and Human Development, Universidade La Salle Canoas, Canoas, Brazil

COVID-19 brought scenes from sci-fi movies into real life. Infected individuals include asymptomatic cases to severe disease leading to death, suggesting the involvement of the genetic constitution of populations and pathogens contributing to differential individuals' outcomes. To investigate shared immunogenic features between SARS-CoV-2 targets and other coronaviruses, we modeled their peptides in 3D structures of HLA-A*02:01 (pMHC), comparing their molecular surfaces. These structures were also compared with a panel of epitopes from unrelated viruses, looking for potential triggers conferring cross-protection in uninfected individuals. As expected, SARS-CoV 1 and 2 peptides share molecular and physicochemical features, providing an explanation for the verified experimental immunogenicity among them. Surprisingly, even discordant sequences from human coronaviruses 229E, OC43 and epitopes from unrelated viruses involved in endemic human infections exhibit similar fingerprints of immunogenicity with SARS-CoV-2 peptides. The same approach indicates a conserved CD8+ T cell recognition between Wuhan SARS-CoV-2 sequences and altered peptides from Variants of Concern. Examination of structural data over epitope sequence analysis here could explain how previous infections may produce a heterologous immunity response in a global scale against emergent diseases such as Covid-19, mitigating its full lethal potential, and paves the way for the development of wide spectrum vaccine development.

KEYWORDS

T-cell epitopes, structural analysis, cross-reactivity, cellular response, immunoinformatics, SARS-CoV-2, Coronaviruses

Introduction

COVID-19 has brought scenes from sci-fi movies into real life. A new virus that spreads in a worldwide fashion, leaving a trail of deaths and fear of the unknown, especially in the beginning. The understanding of all involved aspects in SARS-CoV-2 pathogenesis and mechanisms of immune system elicitation is a crucial step to fighting out the current pandemics and preventing future tragedies. It is important to recognize the SARS-CoV-2 particularities, which can be initially achieved by the comparison with its relatives. The coronaviridae family includes other respiratory syndrome causative agents in humans, like the SARS-CoV and MERS-CoV viruses (1).

In the current pandemic, a growing amount of evidence demonstrates the importance of cellular responses in the SARS-CoV-2 viral clearance (2). Chen et al. (3) presented several works where severe patients exhibited lymphopenia, especially for CD8+ T cells, compared with mild patients or healthy controls. Besides, the cytotoxic T cells in individuals with severe symptoms have diminished levels of cytokine secretion, as YFN-gamma. In order to properly clear the SARS-CoV-2 infection, both arms of immune response must be triggered, and act in a coordinated way, to properly clear the SARS-CoV-2 infection (4–6). So, in addition to the production of neutralizing antibodies, CD4+ and CD8+ T cell responses should be activated (7, 8). These elements include a set of proteins, such as proteasome subunits, TAP1/2 and MHC, orchestrating cellular immune responses as central players. Most of their genes belong to the MHC locus, the most polymorphic human genomic region (9), and this variability enables animal species to face the mutational potential of viruses. Such an effect can be more evident at a populational level. Briefly, these proteins are involved with the processing and presentation of small peptides in cell membranes, allowing the immune system the discrimination of self from non-self, thus eliminating pathogens and tumors. At this point, a critical question arises: not only the MHC genes are involved in the immune response, but also other proteins (TAP, proteasome subunits, and proteins from Peptide Loading Complex), fundamental to producing true epitopes, belong to this genomic region. Many works, aiming to prospect tumoral or vaccine targets, focus their predictions only on the ligandome from proteins of pathogens or cancer samples (10). A potential to bind to different MHC alleles does not confer to peptides their full potential to be a T cell epitope. The complete triggering of a T cell synopsis demands additional requirements, involving epitope immunodominance and pMHC : TCR physicochemical complementarity. Thus, *in silico* analysis considering additional steps on the antigen processing pathway and comparisons among putative targets and immunogenic epitopes, could present a better performance to prospect actual T cell epitopes, as in the current situation where no previous information is available.

It is known that different HLA alleles can bind and present the same viral target (peptides) with altered efficiencies, which could provoke both susceptibility or resistance to a disease caused by a specific pathogen, depending on the type of MHC that the individual possesses (11). Thus, some HLA allotypes are unable to present some recognized immunodominant epitopes, which could impair the efficient immune response triggering. The opposite also occurs, the existence of HLAs with improved potential to present optimal viral targets, allowing infection control (12). The work of Agerer et al. (13) already demonstrates that SARS-CoV-2 mutations in MHC-I-restricted epitopes could evade CD8+ T cell responses by altering the anchor residues that provide stability. It is important to notice that not only the MHC binding is important to immunogenicity, but also the region of interaction with TCR in pMHC complexes provides information able to elicit a T cell effector response, with a percentile of these cells that will constitute the memory cells reservoir. In this sense, it is interesting to investigate immunogenic footprints left by previous infections, searching for cross-reactive targets with SARS-CoV-2 epitopes. Probably, those targets are recognized by the memory T cells pool from the world population individuals', which could be granting a buffering effect, avoiding a more lethal pandemic. The importance of cross-reactivity in Covid-19 is discussed in the work of Saletti et al. (14). They observed that older adults lack SARS-CoV-2 cross-reactive T lymphocytes directed to human coronaviruses, OC43 and NL63, which could be a partial explanation for the more severe clinical outcome observed in this group.

HLA-A*02:01 as an example of universal protective allele in human populations

We hypothesized that pre-existing T cell responses against SARS-CoV-2 are through prevalent alleles in populations, such as the HLA-A*02 supertype. Migliorini et al. (2021) revised the literature and found some works where HLA-A*02:01 was associated with a risk of infection and worse prognosis in some populations. Initially, it seems to be contradictory reasoning, but considering mortality rates in other SARS causative viruses (around 10% in SARS-CoV-1 and 35% in MERS) (15, 16), the current values (slightly above 2%) may indicate a more effective worldwide cellular response. The idea of using HLA-A*02:01 as a model is supported by many studies reporting it as a pivotal allele in viral infections. A consultation on Immune Epitope Database (www.iedb.org) returned 896 references describing positive T cell response against viral peptides presented in the context of the HLA-A*02:01 allele (accessed on March 6, 2021). Besides, 152 out of 1146 immunogenic epitopes described for coronaviruses are restricted to HLA-A*02:01. During the SARS outbreak of

2003, lymphocytes from previously infected individuals were able to eliminate cells presenting SARS-CoV epitopes restricted to HLA-A*02:01 molecules up to six years later from recovering (17–19), demonstrating the importance of this allele on viral clearance and T cell central memory. A conducted study with individuals from China and Hong Kong evidenced that more than half of the SARS recovered subjects were HLA-A*02:01 positive (20). In the present pandemic Italian individuals harboring haplotypes containing the HLA-A*02:01 allele presented a negative correlation between incidence and death (21). In another work, a risk score was developed by Shkurnikov et al., 2021 aiming to predict the possibility of an individual to present severe acute respiratory syndrome coronavirus 2, associated with the presence of the HLA-A*02:01 allele with low risk (22). In Spanish patients, this allele was overrepresented in mild patients, suggesting its involvement in a better prognosis of the disease (23, 24). It is noteworthy considering its high-prevalence among human populations, with frequencies among 57 analyzed countries presenting a median of about 17.7 (Figure 1). While we cannot assign its ability to confer protection, independently, by the above-presented value, this ubiquitous feature enables it as one of the main tools in the proposed global shield.

In our analysis, most of the recovered HLA-A*02:01 epitopes described in past epidemics seem to present conserved sequences compared to their equivalent in SARS-CoV-2 (Table 1), as described in other works (25, 26). Nevertheless, in the case of discordant peptide sequences, the simple sequence comparison of SARS-CoV-1 and SARS-CoV-2 may provide us with little

information about the impact of these alterations concerning the immunogenic potential on current putative viral targets. We could not state that the immunogenicity is preserved or lost, just looking at these linear alterations. Stervbo et al., suggested that previous immunity to specific SARS-CoV-2 proteins is not driven by near-identical epitopes (27). Thus, the analysis of structural and physicochemical features in the peptide-MHC (pMHC) surfaces that contact the T cell receptors may provide us with additional information involved in T cell activation. Such structural investigation has already demonstrated its potential, explaining differential immunogenicity among epitopes from diverse viral strains or tumoral origins (28, 29). Since there is no available crystal for all specific targets (peptides) complexed in HLA-A*02:01 allele, we constructed all customized complexes through our reliable DockTope tool for pMHC modeling (<http://tools.iedb.org/docktope/>) (30). This structural analysis gives us two alternative scenarios: the TCR interacting surfaces of both pHLA-A*02:01 complexes were similar (Figure 2A), explaining the preserved immunogenicity observed in 19 out of 20 SARS-CoV-2 discordant sequence targets (except for ILPDPSKPS, with no positive T cell assay in SARS-CoV-2 – Table 1) (Figure S1); or the targets presented subtle physicochemical alterations in complexes harboring SARS-CoV-2 peptides, compared to former SARS-CoV. Amazingly, in this case, some of them turned into closely related surfaces from previously described immunogenic epitopes from non-related viruses (Figure 2B). In the depicted example, the SARS-CoV-2 SIIAYTMSL peptide demonstrates structural convergence of physicochemical features with the immunodominant Influenza virus epitope

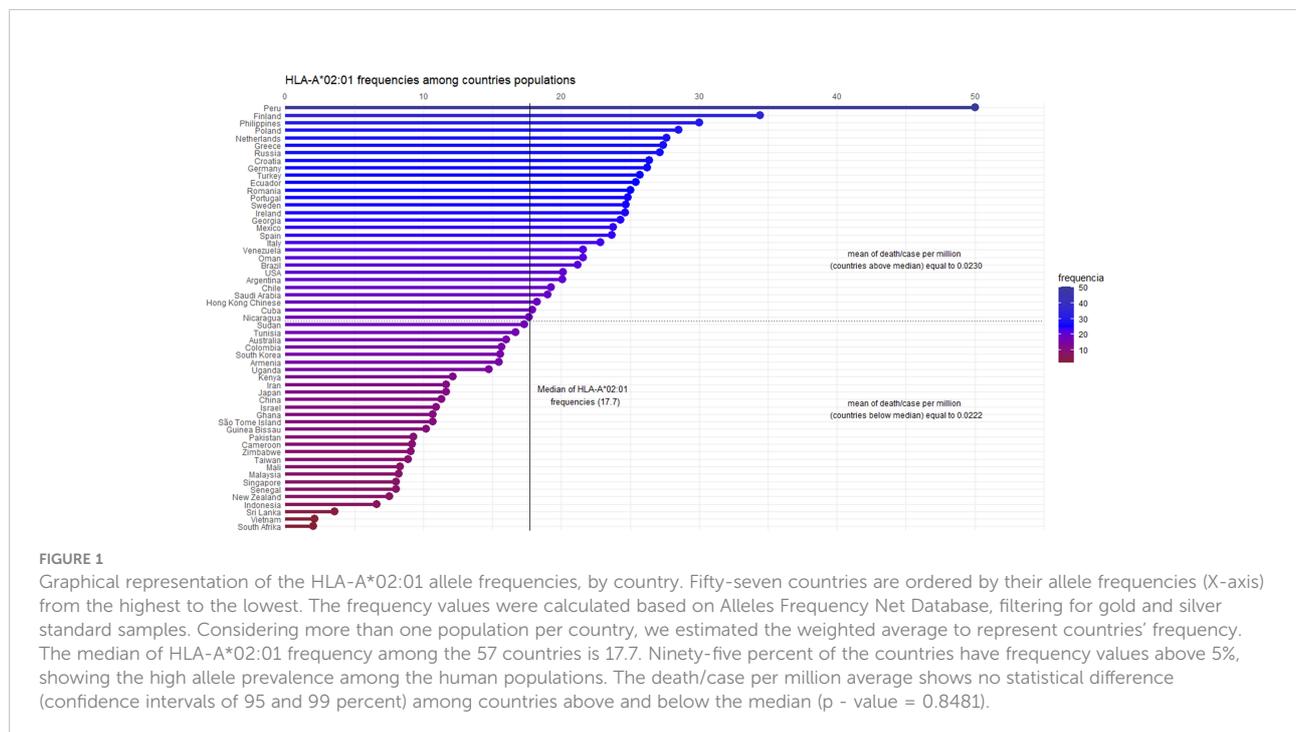


FIGURE 1

Graphical representation of the HLA-A*02:01 allele frequencies, by country. Fifty-seven countries are ordered by their allele frequencies (X-axis) from the highest to the lowest. The frequency values were calculated based on Alleles Frequency Net Database, filtering for gold and silver standard samples. Considering more than one population per country, we estimated the weighted average to represent countries' frequency. The median of HLA-A*02:01 frequency among the 57 countries is 17.7. Ninety-five percent of the countries have frequency values above 5%, showing the high allele prevalence among the human populations. The death/case per million average shows no statistical difference (confidence intervals of 95 and 99 percent) among countries above and below the median (p - value = 0.8481).

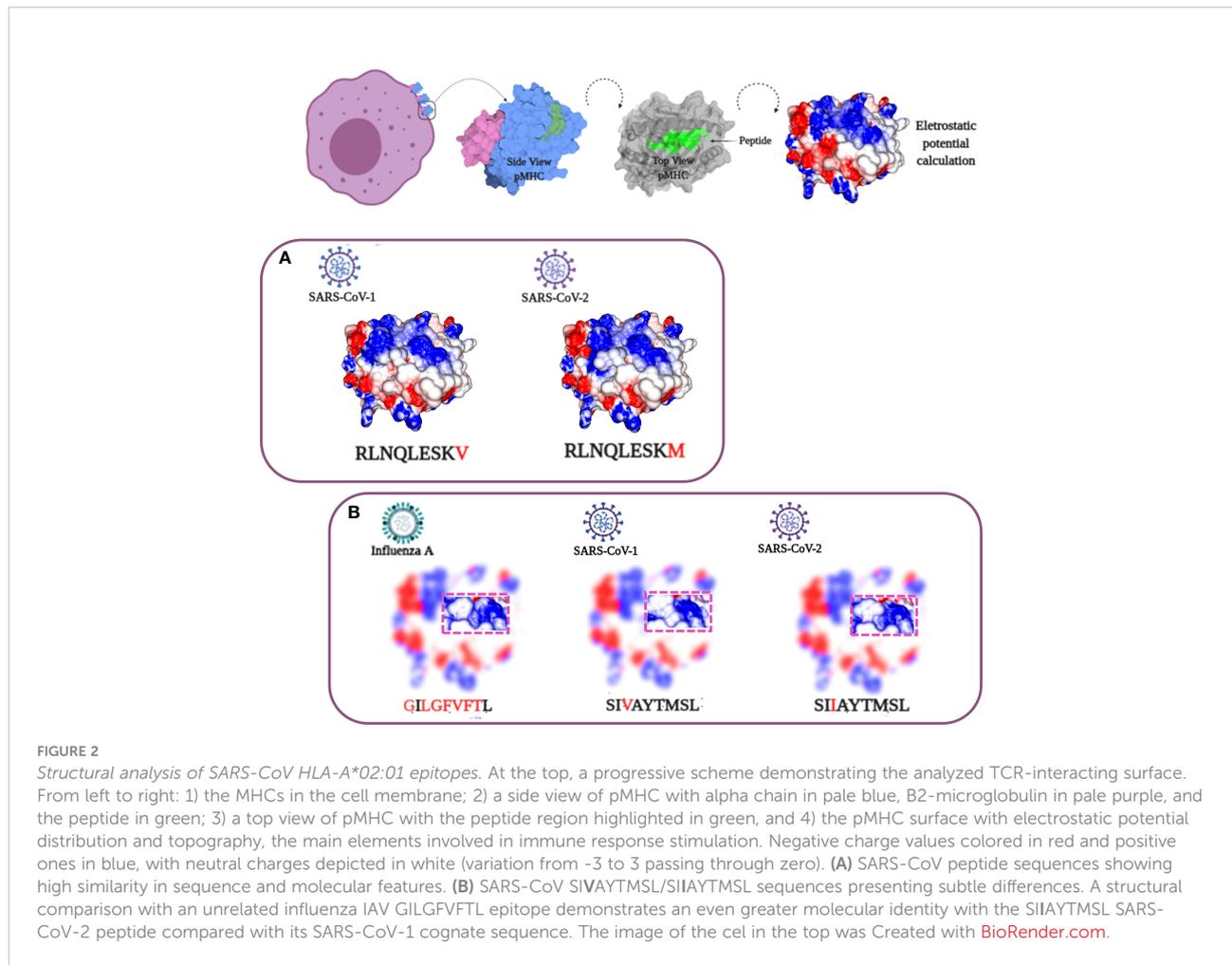
TABLE 1 Recovered HLA-A*02:01 epitopes from SARS-CoV SARS-Cov-2 and other alpha and beta coronaviruses members.

IEDB - experimental positive epitopes from coronaviruses ^A					Corresponding peptides in other coronaviruses ^B			
Epitope ID	Description	Antigen Name	Organism Name	Experimental Assays (Positive/All)	SARS-CoV-2 ^C	HCoV-229E	HCoV-NL63	HCoV-OC43
2801	ALNTLVKQL	S protein	SARS-related coronavirus	2/2		SLNHLTSQL	ALNHLTSQL	ALNNLLQQL
2802	ALNTPKDHI	Nucleoprotein	SARS-related coronavirus	2/2		RVTVPKDHP		DVNTPADIV
16156	FIAGLIAIV	Spike glycoprotein precursor	SARS-related coronavirus	2/2				FINGIFAKV
21347	GMSRIGMEV	Nucleoprotein	SARS-related coronavirus	10/10				
27182	ILLNKHIDA	Nucleoprotein	SARS-related coronavirus	1/3				
27241	ILPDPLKPT	Spike glycoprotein precursor	SARS-related coronavirus	2/2	ILPDPSKPS			
34851	LALLLLDRL	Nucleoprotein	SARS-related coronavirus	4/4				
36724	LITGRLQSL	Spike glycoprotein precursor	SARS-related coronavirus	5/8		LITGRLAAL	LITGRLAAL	LINGRLTAL
37473	LLDRLNQL	Nucleoprotein	SARS-related coronavirus	11/12				
38881	LQLPQGTTL	Nucleoprotein	SARS-related coronavirus	4/6		QKLPNGVTV		GTVLPQGYV
44814	NLNESLIDL	S protein	SARS-related coronavirus	3/4		NINSTLVDL		VLNHSYINL
54690	RLNQLESKV	Nucleoprotein	SARS-related coronavirus	9/11	RLNQLESKM			
58730	SIVAYTMSL	S protein	SARS-related coronavirus	3/3	SIIAYTMSL			
69657	VLNDILSRL	S protein	SARS-related coronavirus	3/4			ETNDVSSML	SLQEILSRL
71663	VVFLHVITYV	Spike glycoprotein precursor	SARS-related coronavirus	5/5				LYFIHENYV
125100	ILLNKHID	Nucleoprotein	SARS-related coronavirus	1/1				
21041	GLMWLSYFV	Membrane glycoprotein	SARS coronavirus TJF	7/8	GLMWLSYFI	LVMWVMYFA	LCLWVMYFV	IIMWIVYFV
64710	TLACFVLAIV	Membrane glycoprotein	SARS coronavirus TJF	5/5			VLALSIFDCFV	
32069	KLPDDFMGCV	Spike glycoprotein precursor	SARS coronavirus BJ01	2/2	KLPDDFTGCV			YSFDSYLGCV
54680	RLNEVAKNL	Spike glycoprotein precursor	SARS coronavirus BJ01	1/1		RLNYVALQT	TLQEFAQNL	RLQEAIKVL

^AThe selection and data of the epitopes were accessed on Immune Epitope Database and Analysis Resource (IEDB - <https://www.iedb.org/> - accessed on March 2021).

^BComparative scheme among the selected positive epitopes from coronaviruses SARS-CoV-2 peptide and other coronaviruses. The blank cells in the columns mean there is no amino acid difference among SARS-CoV-1 epitope and its equivalent in the depicted viruses. Discrepant amino acids were bolded discrepancies were **bolded**.

^CAll corresponding peptides in SARS-CoV-2 were already confirmed in T cell assays except the peptide sequence ILPDPSKPS.



M1₅₈₋₆₆ GILGFVFTL. Both epitopes share 2/9 amino acids, reinforcing the importance of structural investigation over sequence comparison to prospect cross-reactive targets. An impressive point is that other work using analysis of specific TCR sequences for SARS-CoV-2 had a known specificity for the Influenza virus M1 immunodominant epitope. Here, we present a structural basis for it (31). These first comparisons uncovered exciting scenarios. Firstly, even those peptides with sequence alterations in SARS-CoV-2, but without molecular modifications in the TCR interacting surfaces, reveals that our method is reliable to prospect good candidates for immunization strategy in new pandemic events. Moreover, some of these peptides' sequences resemble highly immunogenic epitopes from other viral organisms, which emphasizes a necessity to investigate this new face of the immunogenic prism, that is, previous infections triggering cross-reactivity events. In our analysis, to highlight that the comparisons were not resulting from structural biases, we provide a small sample of TCR interacting surfaces from immunogenic epitopes restricted to HLA-A*02:01, from CrossTope database (www.crosstope.com) (Figure S2) (32).

Searching for SARS-CoV-2 shared immunogenic fingerprints in targets from HCoVs and other prevalent viruses in human populations

The observed molecular similarity between pMHCs complexes containing peptides from SARS-CoV-2 and Influenza virus brings us to another attractive hypothesis that refers to a universal previous cytotoxic response present in populations from all over the world, triggered by previous infections. The first suspects in the investigation were past contact with targets from remaining betacoronavirus (OC43) and alphacoronavirus genus members (229E and NL63). Epidemiological studies reported that 15-30% of the common cold events are caused by this group of pathogens (33). Even considering that they are viruses with a zoonotic origin, we would expect many spillover events throughout the history of humans, maintaining regular contact with our species (34). A codon usage analysis, involving BCoV and HCoV-OC43, suggests that an ancestor coronavirus could be present even 200

kyr ago, in early people (35). Therefore, we would expect that this group of pathogens has also contributed to shaping our current immune system repertoire. Guided by this supposition, we compared the immunogenic SARS-CoV epitopes with 229E, OC43, and NL63 corresponding protein sequences, looking for shared elements involved in immunogenicity triggering. Such analysis presented a clear example where sequence comparison might be hiding shared patterns not detectable by single amino acid identity alignment. In **Table 1**, the sequence identities ranged around 50%, a value usually not reaching thresholds of detection by regular alignment methods prospectation. Nevertheless, when we inspect these same epitopes in the context of pMHC structural models harboring these peptides sequences from SARS-CoV-1, SARS-CoV-2, and alpha and betacoronavirus in HLA-A*02:01 alleles, an intriguing fingerprint arose. A similar electrostatic distribution and topography on the TCR interacting surfaces from the pMHCs, can be observed among SARS-1, SARS-2, and other coronaviruses members (229E and OC43) (**Figure 3A**). It is important to reinforce that both 229E and OC43 putative epitopes were predicted as strong binders to HLA-A*02:01 (data not shown), strengthening their potential as actual triggers for SARS-CoV cross-reactivity. The peptides from the SARS-CoV-1 and 2 are very similar (**GLMWLSYFL** and **GLMWLSYFV**). However, the corresponding peptides from 229E (**LVMWVMYFA**) and OC43 (**IIMWIVYFV**) are pretty divergent (4/9 compared with the same SARS peptides),

evidencing the importance of the structural investigation. Furthermore, other peptides derived from alpha-CoV viruses presented a less prominent but still interesting similarity with immunogenic SARS-CoV-1 epitopes (data not shown). However, they are also potential targets to investigate. A work recently showed that 35% SARS-CoV-2 of seronegative healthy individuals presented S-reactive CD4+ T cells. These cells react almost exclusively with the C-term epitopes region, characterized by higher similarity with spike protein of human endemic common cold coronaviruses. Nevertheless, none of the putative cross-reactive epitopes are pointed out, nor the structural basis hypothesized, reinforcing our propositions (36). Evidence of many CD4+/CD8+ cross responses against many SARS-CoV proteins in unexposed individuals was extensively described in Griffoni et al. (2020), without the specific identification of sequence targets. Other work describing correlations of CD4+/CD8+ T cell differential phenotypes between acute (highly activated cytotoxic) and convalescent (stem-like memory) patients was conducted by Karolinska COVID-19 Study Group (37). Interestingly, they described the occurrence of SARS-CoV-2-specific T cell responses elicited in the absence of circulating antibodies in non-infected individuals, suggesting that previous contacts could be the triggers of these cross-reactive events.

In this regard, given the previous identification of a similar target from the heterologous Influenza virus (M158-66 GILGFVFTL) with a SARS-CoV epitope, the next step was to

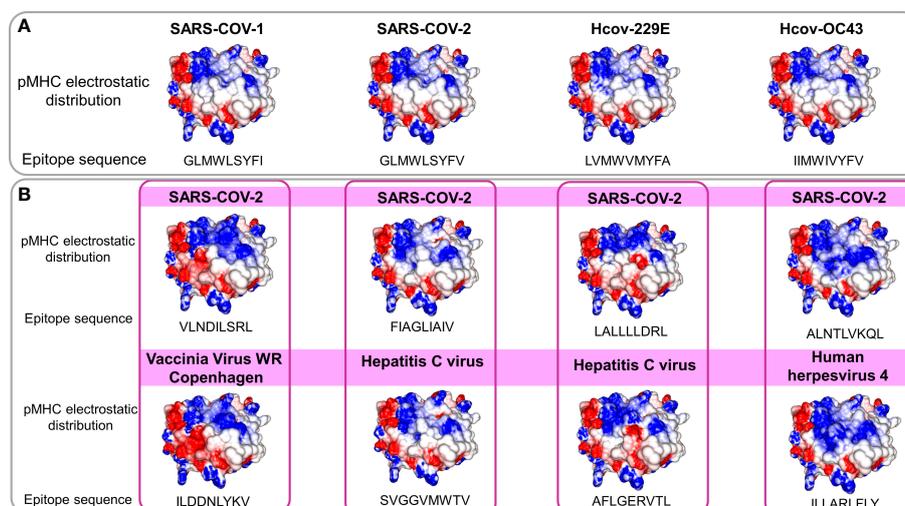


FIGURE 3

Comparison of electrostatic surface distribution and topography of SARS-CoV peptides with alphacoronaviruses and other prevalent viruses in human populations. pMHCs models compared in terms of topography and electrostatic distribution. In **(A)**, two beta (SARS-CoV-1 and 2) and two alpha and beta (229E and OC43), coronaviruses show shared electrostatic distribution and topography despite their sequence divergences. In **(B)**, a panel of four SARS-CoV peptides (top) compared with pMHCs containing viral epitopes from members of other viral families (bottom). Disregarding the lack of sequence identity and phylogenetic relationship, models present structural similarities. Electrostatic calculations are depicted as negative (red) and positive (blue) charges.

scrutinize other trigger candidates on previously described viral epitopes. To perform the comparison, we recover pMHC structures on CrossTope T cell epitope databases (<http://crosstope.com/>) searching for immunogenic fingerprints common to SARS-CoV-2 epitopes and unrelated viruses. The results from these comparisons were newsworthy. When we look for pHLA-A*02:01 structures, not only the previous example of M1₅₈₋₆₆ IAV epitopes matched with SARS epitopes. Ten out of twenty CD8+ coronavirus epitopes, recovered from Immune Epitope Database, have counterparts in targets from common circulating viruses, concerning depicted molecular features. **Figure 3B** presents four examples of SARS-CoV peptides presenting stunning structural identity with unrelated viral epitopes. The matched targets belong to viruses from three different families (Herpesviridae, Poxviridae, and Flaviviridae), being important to show that these targets would probably not be investigated and selected in an approach using conventional methods, given that no apparent identity is presented by any of these structurally related epitopes with SARS-CoV sequences. The remaining comparisons can be viewed in **Figure S3**. Importantly, when we consider these image correspondences, the comparisons with pMHCs from unrelated viruses were more conspicuous than those from other representatives of HCoVs. It seems a paradox, given the natural expectation (considering its phylogenetics proximity) of a more intimate relation between alpha and beta CoVs peptides with SARS-CoV epitopes. In the future, maybe we need to change our prospection point of view to a more general search not only restricting it to phylogenetically related members.

The impact of variants of concern on T cell epitope recognition

The arising of Variants of Concern (VOC) became prevalent in the last months, increasing the number of infections and presenting a more pronounced contagious potential, which probably occurs due to their mutational landscape favoring an increased ACE-2 receptor affinity (38, 39). Some lineages such as B.1.1.7 were associated with the highest in-hospital mortality, which was 20% higher in the second wave of infection (40). Nevertheless, it was caused by an increased demand concomitant with a lack of structure to properly care for these patients, especially in developing countries. VOC infection provokes similar clinical manifestations as the wild-type strain, however, a cohort study from Challen (41) evidenced an increased virulence in the UK variant, which should be confirmed in studies conducted during warmer seasons to exclude this variable in the enhanced risk of death caused by this variant. Two recent works did not present an increase in the severity and number of the symptoms of deaths in patients infected with B.1.1.7 lineage (42, 43). A contributing aspect to be investigated refers to the initial inoculum, which could be amplified by the S protein gain of affinity, resulting in a viral load augmentation, with higher disease severity and death rates increment (44, 45). It seems that even higher infectivity is

counterbalanced by a preserved cellular response. A recent study conducted by researchers of La Jolla Institute demonstrated that epitope mutations do not disrupt CD4+ and CD8+ T cell responses (25). The present work provides some examples of structural T cell recognition sustenance, which could explain the response maintenance in individuals infected by the ancient strain or those vaccinated with Pfizer and Moderna vaccines (tested in La Jolla Institute study), immunogens that stimulate endogenous production of antigen, facilitating its presentation to T cells. We investigated if some of the mutations presented by VOCs include the sequence positions of our analyzed epitopes. Four variants have amino acid alterations in immunogenic epitopes corresponding sequence regions (**Figure 4**). In all cases, there are no substantial physicochemical alterations that could be abolishing their TCR cross-recognition. It emphasizes the cross-protection potential of wild types and vaccines favoring an antigen processing pathway of their antigens against the VOCs. Substitutions in epitope viral strains are responsible for abolishing the response in the Hepatitis C virus. In this case, the mutations cause an alteration in the electrostatic potential distribution of wild-type and the variant strains (**Figure S4**), which was not observed in SARS-CoV-2 VOC examples. In an experiment performed by Nesterenko et al., 2021 (46) epitope homologs were identified and synthesized and T cell cross-reactivity was assessed *via* peptide titration assay. These different coronaviruses targets exhibited a diverse pattern of reactivity with different degrees of variation in the amino acid sequence, which could not be explained by sequence analysis. When we modeled the epitopes from different alpha and beta coronaviruses members, the cross-reactive peptides presented the more similar pMHC surfaces (OC43 and SARS-CoV-2) (**Figure 5**).

Conclusion

In COVID-19, the aetiological agent, SARS-CoV-2, is well established. Nevertheless, the same cannot be deep-stated regarding all elements participating in severe acute respiratory syndrome (SARS) etiology. For this purpose, aspects of cellular response were approached: the structural analysis of immunogenic features presented by SARS-CoV-2, its variants, and unrelated viral epitopes. The epitope structural analysis and their relationship with other HCoVs and unrelated viral targets unveiled noteworthy observations. These pieces of evidence may open two avenues of investigation.

The high degree of molecular conservation between SARS-CoV-1 epitopes and their corresponding sequences in SARS-CoV-2 allows its use in vaccine development to stimulate cross-reactive responses, covering distinct SARS-CoV-2 strains, for example. In this regard, animal reservoirs should be inspected looking for beta CoVs with the potential to spill out of their natural hosts to humans. Peptides sequences from these putative HCoVs pathogens could be structurally compared searching for cross-reactive T cell targets to be used in a virtual future

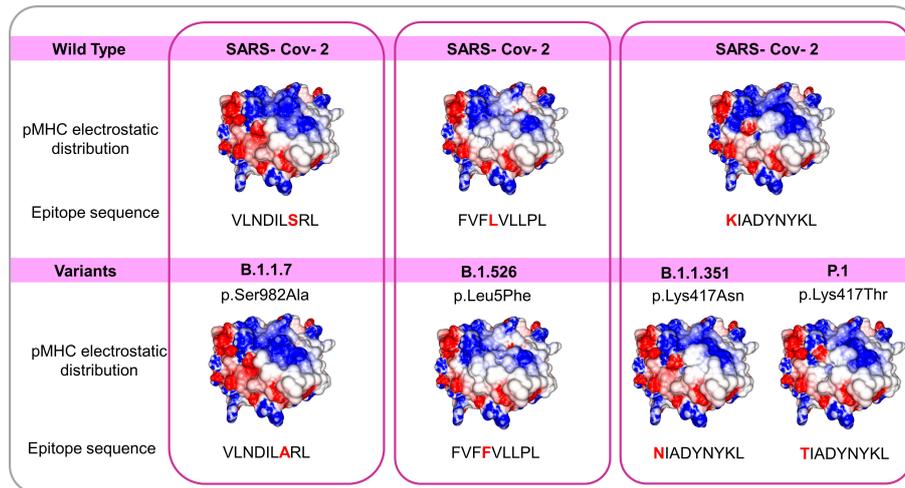


FIGURE 4
 Comparisons among pMHCs carrying wild SARS-CoV-2 epitopes and the correspondent variant sequences in SARS-CoV-2 Spike protein. The electrostatic distributions surfaces from the wild type and correspondent variants do not differ, even presenting amino acid changes in peptides suggesting a preserved TCR recognition. The variants B.1.1.7 B.1.526 B.1.1.351 and P.1 with their respective changes (S982A L5F K417N and K417T respectively) and correspondent epitope sequences.

occurrence of a new coronavirus spillover phenomenon. Such preventive strategies could abbreviate steps to develop immunotherapeutic methods, avoiding the emergence of new pandemics. The second line of the investigation resulted in an even more attractive hypothesis. Previous infections with different alpha/beta-CoVs and unrelated common viruses can be generating memory T cells against SARS-CoV-2 in a significant portion of the population. This pool of cells in different individuals could be providing a universal immunogenic shield against SARS-CoV-2 and, probably,

against other potentially emergent and endemic viruses. This mechanism seems to be evolutionarily constructed by regular cross-reactive contacts. Moreover, the defense appears to be associated with prevalent alleles, which probably present peptides harboring fingerprints of immunogenicity shared by epitopes that regularly infect humans. The last investigated point was the VOIs substitutions present in immunogenic regions of SARS-COV-2 epitopes. We observed that the mutations do not alter the TCR interaction region of pMHCs. In this way, it seems that the CD4+ and CD8+ responses against these targets are

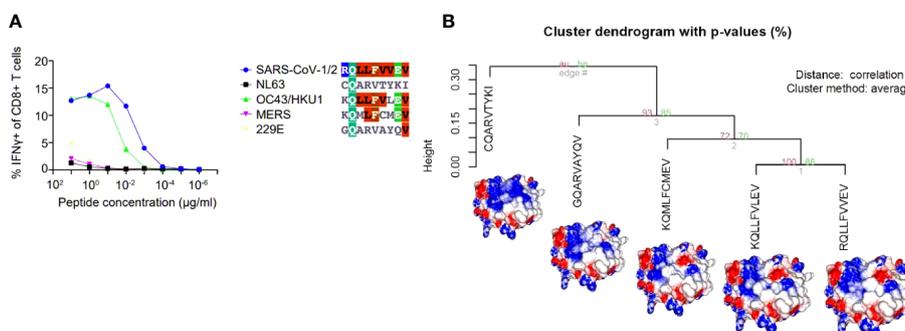


FIGURE 5
 A hierarchical clusterization of pMHC's electrostatic surfaces of different epitope sequences belonging to distinct coronaviruses (SARS-CoV-1/2, NL63, OC43/HKU1, MERS, 229E). In the chosen example, five sequences (in the context of HLA-A*02:01) from the analysed coronavirus were challenged against the same TCR and presented the pattern of recognition showed in (A). In (B), we depicted the HCA and images from the customized pMHC model complexes demonstrating that the epitopes that elicited the cross-reactive response were the most similar (from OC43/HKU-1 and SARS-CoV-1/2), providing an explanation for the observed pattern of cross-reaction. The figure presenting the results in A was extracted, and modified from (46).

preserved, where the elicited T cells responses seem to be directed toward constrained regions of these viruses.

Methods and resources

HLA-A*02:01 allele frequencies among the countries

The HLA-A*02:01 distribution and its frequencies for all included countries were obtained from The Allele Frequency Net Database (47). The applied filter includes only gold and silver standard population samples, considering $n > 50$ and random studies. Weighted frequencies were calculated based on all populations' frequency information for the resulting 57 countries. Table S1 presents the frequencies and sample sizes (total population) for each country. The `dplyr` and `ggplot2` packages in the RStudio evaluate the distribution of data, represented in Figure 1.

SARS-CoV Epitopes prospectings

The Immune Epitope Database and Analysis Resource (IEDB - <https://www.iedb.org/>) (48) was accessed to prospect and recover experimental coronavirus epitopes (the search was refined through the SARS-CoV-1 organism, T cell positive assays, and MHC allele), resulting in 20 immunogenic peptides restricted to HLA-A*02:01 allele (Table 1 compiles its Epitope IDs). Those epitopes derived from the N (nucleocapsid) protein, Surface (spike) protein, and Membrane glycoprotein. Aiming to check if they have a counterpart in the SARS-CoV-2 proteome, we perform a Needleman-Wunsch Global Align Nucleotide Sequences (BLAST) using the protein sequences from 2002/2003 SARS virus against SARS-CoV-2 proteome sequence from Wuhan ancient genome. NetMHCcons tool (49) was applied to predict binding affinity in the discordant SARS-CoV-2 sequences and also for other HCoV peptides. The 229E and NL63 Alphacoronaviruses plus the OC43 Betacoronavirus strains were also checked for corresponding SARS-CoV epitopes sequences, screening their protein correspondences searching for some amino acid identity with immunogenic targets described for SARS-CoV-1.

Variants prospection

The panel of amino acids reported changes in the spike (S) protein in SARS-CoV-2 VOCs (50–52) were screened to evaluate if they fall in the HLA-A*02:01 epitopes included in our analysis. Four changes were matched to three epitopes: B.1.526 (p.Leu5Phe) to epitope FVFLVLLPL, B.1.1.351 (p.Lys417Asn) and P.1 (p.Lys417Thr) to epitope KIADYNYKL, and B.1.1.7 (p.Ser3

82Ala) variant occur in the epitope VLNDILSRL already described in Table 1. Spike protein sequences were verified in UniProt code: P0DTC2 (SPIKE_SARS2).

Structural analysis

All the recovered epitopes for Alpha and Beta Coronaviruses and analysed VOCs were modeled in HLA-A*02:01 context to investigate their TCR interaction surfaces looking for immunogenicity fingerprints and shared structural features. The full applied rationale can be found in (53). The Docktope tool (<http://tools.iedb.org/docktope/>) allowed to produce pMHC models containing the analysed peptides anchored to the HLA-A*02:01 Class I MHC (pMHC). This tool allowed the generation of models based on a D1-EM-D2 approach, involving a first step of molecular docking (D1) performed using Autodock Vina program between the MHC-I and the epitope provided; followed by an energy minimization step (EM) to correct possible steric clashes between the epitope and the MHC-I; and a second round of molecular docking (D2) aiming to refine the structure.

The generated tridimensional structures were inputted in GRASP to calculate their electrostatic surfaces, verifying amino acid changes impact on the overall charge distribution in those models. The GRASP application allowed to generate colors according to the charge distribution considering the model surface: the electrostatic potential ranges from -5 (negative charges in red) to 5 (positive charges in blue) passing through neutral (white color). To quantify this investigation, we utilize the software ImageJ to analyze the RGB content of several spots in the pMHC structure where contact with the TCR tends to occur and save this data in the form of mean, mode, and standard deviation in a numeric form. The numeric RGB data is then clusterized utilizing the R software package Pvcust to perform the hierarchical clusterization based on the data provided and separate different structures from each other while clustering similar structures based on their electric charge distribution.

Images generated for those complexes (pMHC) were compared against all structures included at the CrossTope Structural Databank, looking for shared patterns which may indicate cross reactivity among structurally similar epitopes. The CrossTope database contain pMHC images of distinct immunogenic epitopes described in the literature for prevalent human alleles. The dataset images were analysed utilizing an electrostatic range which varies from -5 to 5.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Authors contributions

EA: Visual and Statistical Analysis, Investigation, Formal analysis. MB: Formal analysis, Resources, MM: Data curation, Investigation, Writing. GV: Conceptualization, Writing, Supervision, Project administration. All authors contributed to the article and approved the submitted version.

Funding

This work was supported by scholarships from National Council for Scientific and Technological Development (CNPq); and National Council for Improvement of Higher Education (CAPES).

Acknowledgments

We thank the scholarships from National Council for Scientific and Technological Development (CNPq) and National Council for the Improvement of Higher Education (CAPES) for their support.

Conflict of interest

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

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the

References

1. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature* (2020) 579:265. doi: 10.1038/s41586-020-2008-3
2. Moss P. The T cell immune response against SARS-CoV-2. *Nat Immunol* (2022) 23:186–93. doi: 10.1038/s41590-021-01122-w
3. Chen Z, John Wherry E. T Cell responses in patients with COVID-19. *Nat Rev Immunol* (2020) 20:529. doi: 10.1038/s41577-020-0402-6
4. Shi Y, Zhou G, Li Q. Asynchronous actions of immune responses in COVID-19 patients. *Signal Transduct Target Ther* (2020) 5:284. doi: 10.1038/s41392-020-00424-z
5. Sewell HF, Agius RM, Stewart M, Kendrick D. Cellular immune responses to covid-19. *BMJ* (2020) 370:m3018. doi: 10.1136/bmj.m3018
6. Rydzynski Moderbacher C, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, et al. Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. *Cell* (2020) 183:996. doi: 10.1016/j.cell.2020.09.038
7. Swadling L, Maini MK. T Cells in COVID-19 - united in diversity. *Nat Immunol* (2020) 21:1307. doi: 10.1038/s41590-020-0798-y
8. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, et al. Broad and strong memory CD4. *Nat Immunol* (2020) 21:1336. doi: 10.1038/s41590-020-0782-6
9. Arango MT, Perricone C, Kivity S, Cipriano E, Ceccarelli F, Valesini G, et al. HLA-DRB1 the notorious gene in the mosaic of autoimmunity. *Immunol Res* (2017) 65:82. doi: 10.1007/s12026-016-8817-7
10. Mösch A, Raffegerst S, Weis M, Schendel DJ, Frishman D. Machine learning for cancer immunotherapies based on epitope recognition by T cell receptors. *Front Genet* (2019) 10. doi: 10.3389/fgene.2019.01141
11. Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* (2013) 14:301. doi: 10.1146/annurev-genom-091212-153455

reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fimmu.2022.931372/full#supplementary-material>

SUPPLEMENTARY FIGURE 1

Comparison of epitope sequences and pMHC electrostatic distribution among SARS-1 and SARS-2. Five out of 20 paired peptides show divergent sequences. Nevertheless, the presence of amino acid changes does not cause alterations in the pMHC electrostatic distribution pattern. This immunogenicity predicted by the sharing of structural features was already experimentally confirmed, considering that all corresponding SARS-CoV-2 stimulates cytotoxic responses. pMHC electrostatic distribution is colored as follows: from -5 (electronegative in red) to +5 (positive in blue) passing through 0 (neutral white). Divergent amino acids detached in bold.

SUPPLEMENTARY FIGURE 2

TCR interacting surfaces from pMHC immunogenic structures. Nine pMHC from HLA-A*02:01 bound to experimental epitopes. We provided this panel to demonstrate the molecular diversity among immunogenic targets.

SUPPLEMENTARY FIGURE 3

Additional comparisons of SARS-CoV epitopes and other immunogenic epitopes from the CrossTope Database. Shared fingerprints in the above-exemplified pairs of structures, indicating that cross-reactivity non-related triggers can be a widespread phenomenon.

SUPPLEMENTARY FIGURE 4

Two distinct Hepatitis C virus epitopes and their pMHCs electrostatic patterns. The variation (detached in red) between the wild-type peptide sequence (CVNGVCWTV) with its variant peptide (TVGDVMWTV) leads to divergent charge distribution at the indicated pMHC surface area (pink square). This variant peptide, from HCV genotype 3, aborts the cellular response in lymphocytes that recognizes the WT peptide. The electrostatic potential distribution is colored as follows: from -5 (electronegative in red) to +5 (positive in blue), passing through 0 (neutral in white).

12. Dronamraju K. *The compatibility gene: How our bodies fight disease, attract others, and define our selves* by Daniel m. Davis, Vol. 90. (2015). p. 343, The Quarterly Review of Biology.
13. Agerer B, Koblishcke M, Gudipati V, Montaña-Gutierrez LF, Smyth M, Popa A, et al. SARS-CoV-2 mutations in MHC-I-restricted epitopes evade CD8. *Sci Immunol* (2021) 6:57. doi: 10.1126/sciimmunol.abg6461
14. Saletti G, Gerlach T, Jansen JM, Molle A, Elbahesh H, Ludlow M, et al. Older adults lack SARS CoV-2 cross-reactive T lymphocytes directed to human coronaviruses OC43 and NL63. *Sci Rep* (2020) 10:21447. doi: 10.1038/s41598-020-78506-9
15. Chen B, Tian EK, He B, Tian L, Han R, Wang S, et al. And Cheng, w. 2020. overview of lethal human coronaviruses. *Signal Transduct Target Ther* (2020) 5 (1):89. doi: 10.1038/s41392-020-0190-2
16. Janice Oh HL, Ken-En Gan S, Bertoletti A, Tan YJ. Understanding the T cell immune response in SARS coronavirus infection. *Emerg Microbes Infect* (2012) 1: e23. doi: 10.1038/emi.2012.26
17. Chen YM, Liang SY, Shih YP, Chen CY, Lee YM, Chang L, et al. Epidemiological and genetic correlates of severe acute respiratory syndrome coronavirus infection in the hospital with the highest nosocomial infection rate in Taiwan in 2003. *J Clin Microbiol* (2006) 44:359. doi: 10.1128/JCM.44.2.359-365.2006
18. Tsao YP, Lin JY, Jan JT, Leng CH, Chu CC, Yang YC, et al. HLA-A*0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. *Biochem Biophys Res Commun* (2006) 344:63. doi: 10.1016/j.bbrc.2006.03.152
19. Channappanavar R, Fett C, Zhao J, Meyerholz DK, Perlman S. Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. *J Virol* (2014) 88:11034. doi: 10.1128/JVI.01505-14
20. Wang Y-D, Sin W-YF, Xu G-B, Yang H-H, Wong T-Y, Pang X-W, et al. T-Cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. *J Virol* (2004) 78:11. doi: 10.1128/JVI.78.11.5612-5618.2004
21. Pisanti S, Deelen J, Gallina AM, Caputo M, Citro M, Abate M, et al. Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of covid-19. *J Transl Med* (2020) 18:352. doi: 10.1186/s12967-020-02515-5
22. Shkurnikov M, Nersisyan S, Jankevicius T, Galatenko A, Gordeev I, Vechorko V, et al. Association of HLA class I genotypes with severity of coronavirus disease-19. *Front Immunol* (2021) 12. doi: 10.3389/fimmu.2021.641900
23. Iturrieta-Zuazo I, Rita CG, García-Soidán A, de Malet Pintos-Fonseca A, Alonso-Alarcón N, Pariente-Rodríguez R, et al. Possible role of HLA class-I genotype in SARS-CoV-2 infection and progression: A pilot study in a cohort of covid-19 Spanish patients. *Clin Immunol* (2020) 219:108572. doi: 10.1016/j.clim.2020.108572
24. Migliorini F, Torsiello E, Spiezia F, Oliva F, Tingart M, Maffulli N. Association between HLA genotypes and COVID-19 susceptibility, severity and progression: a comprehensive review of the literature. *Eur J Med Res* (2021) 26 (1):84. doi: 10.1186/s40001-021-00563-1
25. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. *Cell Host Microbe* (2020) 27:671. doi: 10.1016/j.chom.2020.03.002
26. Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, et al. Human leukocyte antigen susceptibility map for SARS-CoV-2. *J Virol* (2020) 94 (13):JVI.00510. doi: 10.1101/2020.03.22.20040600
27. Stervbo U, Rahmann S, Roch T, Westhoff TH, Babel N. Epitope similarity cannot explain the pre-formed T cell immunity towards structural SARS-CoV-2 proteins. *Sci Rep* (2020) 10:18995. doi: 10.1038/s41598-020-75972-z
28. Antunes DA, Rigo MM, Freitas MV, Mendes MFA, Sinigaglia M, Lizée G, et al. Interpreting T-cell cross-reactivity through structure: Implications for TCR-based cancer immunotherapy. *Front Immunol* (2017) 8. doi: 10.3389/fimmu.2017.01210
29. Mendes MF, Antunes DA, Rigo MM, Sinigaglia M, Vieira GF. Improved structural method for T-cell cross-reactivity prediction. *Mol Immunol* (2015) 67:303. doi: 10.1016/j.molimm.2015.06.017
30. Rigo MM, Antunes DA, de Freitas MV, Mendes MFdA, Meira L, Sinigaglia M, et al. DockTope: a web-based tool for automated pMHC-I modelling. *Sci Rep* (2015) 5:1. doi: 10.1038/srep18413
31. Sidhom J-W, Baras AS. Analysis of SARS-CoV-2 specific T-cell receptors in ImmuneCode reveals cross-reactivity to immunodominant influenza M1 epitope. *bioRxiv* (2020). doi: 10.1101/2020.06.20.160499
32. Sinigaglia M, Antunes DA, Rigo MM, Chies JA, Vieira GF. CrossTope: a curate repository of 3D structures of immunogenic peptide: MHC complexes. *Database (Oxford)* (2013), 2013:bat002. doi: 10.1093/database/bat002
33. Mesel-Lemoine M, Millet J, Vidalain P-O, Law H, Vabret A, Lorin V, et al. A human coronavirus responsible for the common cold massively kills dendritic cells but not monocytes. (2012). doi: 10.1128/JVI.00269-12
34. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. *Int J Biol Sci* (2020) 16:1686. doi: 10.7150/ijbs.45472
35. Brandão PE Universidade de São Paulo S. P., Brazil. Could human coronavirus OC43 have co-evolved with early humans? *Genet Mol Biol* (2018) 41:692. doi: 10.1590/1678-4685-gmb-2017-0192
36. Braun J, Loyal L, Frentsch M, Wendisch D, Georg P, Kurth F, et al. SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* (2020) 587:270. doi: 10.1038/s41586-020-2598-9
37. Sekine T, Perez-Potti A, Rivera-Ballesteros O, Strålin K, Gorin J-B, Olsson A, et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *bioRxiv* (2020). doi: 10.1101/2020.06.29.174888
38. Bayarri-Olmos R, Rosbjerg A, Johnsen LB, Helgstrand C, Bak-Thomsen T, Garred P, et al. The SARS-CoV-2 Y453F mink variant displays a pronounced increase in ACE-2 affinity but does not challenge antibody neutralization. *J Biol Chem* (2021) 296:100536. doi: 10.1016/j.jbc.2021.100536
39. Ramanathan M, Ferguson ID, Miao W, Khavari PA. SARS-CoV-2 B.1.1.7 and B.1.351 spike variants bind human ACE2 with increased affinity. *bioRxiv* (2021). doi: 10.1101/2021.02.22.432359
40. Abdoool Karim SS, de Oliveira T. New SARS-CoV-2 variants — clinical, public health, and vaccine implications. *New Engl J Med* (2021) 384:1866–8. doi: 10.1056/NEJMc2100362
41. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. *BMJ* (2021) 372:n579. doi: 10.1136/bmj.n579
42. Graham MS, Sudre CH, May A, Antonelli M, Murray B, Varsavsky T, et al. Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: an ecological study. *Lancet Public Health* (2021) 6(5): E335–45. doi: 10.1016/S2468-2667(21)00055-4
43. Frampton D, Rampling T, Cross A, Bailey H, Heaney J, Byott M, et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet Infect Dis* (2021) 21(9):P1246–56. doi: 10.1016/S1473-3099(21)00170-5
44. Memoli MJ, Czajkowski L, Reed S, Athota R, Bristol T, Proudfoot K, et al. Validation of the wild-type influenza a human challenge model H1N1pdm18T: an A(H1N1)pdm09 dose-finding investigational new drug study. *Clin Infect Dis* (2015) 60:693. doi: 10.1093/cid/ciu924
45. Fajnzylber J, Regan J, Coxen K, Corry H, Wong C, Rosenthal A, et al. SARS-CoV-2 viral load is associated with increased disease severity and mortality. *Nat Commun* (2020) 11:5493. doi: 10.1038/s41467-020-19057-5
46. Nesterenko PA, McLaughlin J, Tsai BL, Sojo GB, Cheng D, Zhao D, et al. HLA-A*02:01 restricted T cell receptors against the highly conserved SARS-CoV-2 polymerase cross-react with human coronaviruses. *Cell Rep* (2021) 37(13):110167. doi: 10.1016/j.celrep.2021.110167
47. Gonzalez-Galarza FF, Takeshita LY, Santos EJ, Kempson F, Maia MH, da Silva AL, et al. Allele frequency net 2015 update: new features for HLA epitopes, KIR and disease and HLA adverse drug reaction associations. *Nucleic Acids Res* (2015) 43:D784. doi: 10.1093/nar/gku1166
48. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res* (2019) 47(D1): D339–43. doi: 10.1093/nar/gky1006
49. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: A consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* (2012) 64(3):177–86. doi: 10.1007/s00251-011-0579-8
50. Rambaut A, Holmes EC, O'Toole Á, Verity T, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* (2020) 5:1403. doi: 10.1038/s41564-020-0770-5
51. Wang P, Nair MS, Liu L, Iketani S, Luo Y, Guo Y, et al. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *bioRxiv* (2021). doi: 10.1101/2021.01.25.428137
52. West AP, Barnes CO, Yang Z, Bjorkman PJ. SARS-CoV-2 lineage B.1.526 emerging in the New York region detected by software utility created to query the spike mutational landscape. *bioRxiv* (2021). doi: 10.1101/2021.02.14.431043
53. Vianna P, Mendes MFA, Bragatte MA, Ferreira PS, Salzano FM, Bonamino MH, et al. pMHC structural comparisons as a pivotal element to detect and validate T-cell targets for vaccine development and immunotherapy—a new methodological proposal. *Cells* (2019) 8(12):1488. doi: 10.3390/cells8121488