

Design of Peptide Vaccine for COVID19: CD8+ and CD4+ T cell epitopes from SARS-CoV-2 open-reading-frame protein variants.

Simone Parn¹, Gabriel Jabbour², Vincent Nguyenkhoa³, Sivanesan Dakshanamurthy^{4,*}.

¹University of the District of Columbia, Washington, D.C. 20008.

²School of Medicine, Georgetown University Medical Center, Washington DC 20057.

³University of Virginia, Charlottesville, Virginia 22904.

⁴Molecular and Experimental Therapeutic Research in Oncology (METRO) Program, Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington DC 20057.

*Corresponding Author

Dr. Sivanesan Dakshanamurthy, Ph.D, MBA.

e-mail: sd233@georgetown.edu

Phone: (202) 687 2347

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has challenged public health at an unprecedented scale which has led to a dramatic loss of human life worldwide. To design a protective vaccine against SARS-CoV-2, it is necessary to understand which SARS-CoV-2 specific epitopes can elicit a T cell response and provide protection across a broad population. In this study, PLpro and RdRp, two immunogenic non-structural proteins from an immunodominant gene region ORF1ab, as well as ORF3a and ORF9b are identified as potential vaccine targets against SARS-CoV-2. To select top epitopes for vaccine design, we used various clinical properties, such as antigenicity, allergenicity, toxicity and IFN- γ secretion. The analysis of CD8 and CD4 T cell epitopes revealed multiple potential vaccine constructs that cover a high percentage of the world population. We identified 8 immunogenic, antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD8 proteins for nsp3, 4 for nsp12, 11 for ORF3a and 3 for ORF9b that are common across four lineages of variants of concern – B.1.1.7, P.1, B.1.351 and B.1.617.2, which protect 98.12%, 87.08%, 96.07% and 63.8% of the world population, respectively. We also identified variant specific T cell epitopes that could be useful in targeting each variant strain separately. Including the prediction of mouse MHC affinity towards our top CD8 epitopes, our study revealed a total of 3 immunogenic, antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD8 epitopes overlapping with 6 antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD4 epitopes across all four variants of concern which can effectively be utilized in pre-clinical studies. The landscape of SARS-CoV-2 T cell epitopes that we identified can help lead SARS-CoV-2 vaccine development as well as epitope-based peptide vaccine research in the future.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is a novel causative agent of coronavirus disease 2019 (COVID-19) that belongs to the family of *Coronaviridae* and the genus of *Betacoronavirus* [1]. First emerged in Wuhan, China, the virus rapidly spread across the world, challenging all aspects of human life. On March 11th, 2020, the World Health Organization declared the global emergency a pandemic [2]. To date, there has been more than 222 million COVID-19 cases and over 4.5 million deaths worldwide, which indicate the need for a preventative and effective vaccine [3]. In the recent decades, epitope-based vaccine design has gained more insight due to its time- and cost-effective approach [4]. The vaccine approach is centered around immunoinformatics methods and bioinformatics tools that aid in designing CD4+ and CD8+ T cell epitopes against viral antigens. T cells play an important role in the elimination of virus-infected cells, with CD8+ T cells being critical for the clearance of infected cells, and CD4+ T cells for the promotion of specific B cell antibody production and support of CD8+ T cell response. These immune cells have characteristic T cell receptors (TCR) that interact with major histocompatibility complex I and II molecules located on the surface of antigen-presenting cells (APCs), such as dendritic cells, in order to trigger an immune response against a pathogen [5].

The SARS-CoV-2 genome encodes 4 major structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N), and 6 major non-structural proteins, often referred to as open reading frames (ORFs) that carry vital functions for the transmissibility and survival of the virus. ORF1ab, the largest gene region of the SARS-CoV-2 genome, is highly conserved within coronaviruses and it encodes 16 non-structural proteins (nsp1-16) that are cleaved by two polyproteins, pp1a and pp1ab [6]. In this study, PLpro (nsp3) and RdRP (nsp12), were targeted due to their potential ability to provoke the highest immune response within ORF1ab. Nsp3, the largest multi-domain protein produced by coronaviruses, plays a central role in viral replication, as well as nsp12, which functions as RNA polymerase. ORF3a, the largest accessory protein of 275 amino acids in SARS-CoV-2, also has a significant role in virus pathogenesis and T cell reactivity [7, 8, 9].

In addition to ORF1ab and ORF3a proteins, ORF9b, an alternative open reading frame incorporated into the N gene, will be used as another antiviral target in this study. The 97 amino acid long protein is abundant in SARS-CoV-2 and highly conserved in SARS-like coronaviruses, making it a potential vaccine candidate. Besides this, ORF9b plays a critical role in virus-host interactions [10]. Due to its accumulation early in infection and its interaction with TOM70, a mitochondrial protein essential for RNA sensing adaptor MAVS activation, Orf9b suppresses the innate immune response by blocking type I interferon

(IFN-I) [11]. Therefore, we aim to target the interaction site between ORF9b and TOM70 by identifying immunogenic CD4+ and CD8+ T cell epitopes that would aid in vaccine construction.

The emergence of variants of SARS-CoV-2 makes it more challenging to come up with protective vaccine candidates. The spectra of SARS-CoV-2 mutations harbors a wide range of silent and point mutations that impact the structure and function of the proteins, making the virus more transmissible. With the virus constantly changing, some variants of concern (VOC) have been detected, including the B.1.1.7, P.1, B.1.351 and B.1.617.2 lineages that first emerged in the United Kingdom, Brazil, South Africa and India, respectively [12]. Therefore, in this analysis the lineage-defining sequences for each VOC will be utilized to shed light on the most important and prevalent T cell epitopes that would help design the most effective vaccine.

In this study, we aim to identify immunogenic CD4+ and CD8+ T cell epitopes on nsp3, nsp12, ORF3a and ORF9b proteins of SARS-CoV-2 for effective vaccine development. It is hypothesized that CD4+ and CD8+ T cell recognition is largely based on HLA specificity, and binding to multiple HLA allelic variants is important for the amplification of potential immunogenicity. We utilized Immune Epitope Database (IEDB) server to compare several available tools for their accuracy predicting immunogenic T cell epitopes [13]. Predicted T cell epitopes were compared by their match rate to experimentally validated immunogenic SARS-CoV-2 T cell epitopes. This allowed the identification of the best performing tool to predict both MHC-I and MHC-II restricted T-cell epitopes from SARS-CoV-2 proteins with high binding affinity to MHC class I and II alleles. Beyond MHC affinity, epitopes were further refined by predicting immunogenicity, population coverage, antigenicity, allergenicity, toxicity, IFN- γ secretion, half-life, GRAVY (Grand average of hydropathicity) and other amino acid physiochemical properties, to fully maximize the quality and efficiency of the designed vaccine. The 3D structures of the final CD8 T cell epitopes were thereafter predicted and docked with their respective HLA alleles to visualize the peptide-MHC interaction.

After predicting immunogenic CD4 and CD8 T cells, we attempted to identify top immunogenic CD8 epitopes (8-11 aa long) within immunogenic CD4 epitopes (15 aa long) that together could be used to provoke a strong and long-lasting immune response against the four variants of concern of SARS-CoV-2. In order to safely test the predicted vaccine constructs in pre-clinical studies, we also predicted mouse MHC affinity to our top CD8 epitopes. Mouse models are useful for the evaluation of antiviral targets as they share features with humans and two mouse strains - BALB/CJ and C57BL/6 have been reported as good models for SARS-CoV-2 [14, 15].

Materials and Methods

Retrieval of SARS-CoV-2 sequence

The reference protein sequences of SARS-CoV-2 Wuhan isolate were retrieved from NCBI database using accession number NC_045512. The corresponding protein accession numbers are: NCBI: YP_009725299.1 (NSP3), NCBI: YP_009725307.1 (NSP12), NCBI: YP_009724391.1 (ORF3a) and UniProtKB/Swiss-Prot: P0DTD2.1 (ORF9b).

Retrieval of non-structural protein 3 (nsp3) MHC class I experimental epitopes

To assess the validity of the predictions in the datasets, predicted epitopes were compared to experimentally identified SARS-CoV-2 T cell epitopes. Since nsp3 is the largest multi-domain protein in SARS-Cov-2, the highest number of experimental epitopes was identified and used in the validation. A total of 169 high quality experimental CD8+ T cell epitopes were retrieved from SARS-CoV-2 T cell epitope database that were previously reported in multiple studies [7, 8, 16-23] (**Supplementary Table 1**) and tested for T cell reactivity. Additional 3 epitopes – ILLLDQALV, WLMWLIINL and CLEASFNYL were identified and selected as highly homologous to SARS-CoV-1 genome, having identity of 88.89%, 66.67% and 55.56%, respectively [24].

CD8+ T cell epitope tool comparisons

The IEDB.org website was used to predict SARS-CoV-2 CD8+ T cell epitopes of the reference nsp3 protein using 6 different tools: NetMHCpan EL 4.1 (IEDB recommended), NetMHCpan BA 4.1, NetMHCpan EL 4.0, NetMHCpan BA 4.0 [25-29], ANN 4.0 [30-36], and IEDB consensus [37]. Only MHC I alleles with a frequency >1% were selected for further processing. The duplicate-free peptides were ordered according to their available attributes, such as percentile rank and IC50 value, and then matched to the experimental epitopes using three levels of stringency: lenient, intermediate and stringent. With the lenient method, a predicted epitope matches an experimental epitope if the predicted epitope is included within the experimental epitope or vice versa. With the intermediate method, two epitopes match if they only differ from in up to one single-character edit (insertion, deletion or substitution). With the stringent method, a predicted epitope is only considered a match if it shares 100% sequence similarity with the experimental epitope. To select the best tool, ROC curves with respective AUCs were generated to compare each tool-attribute combination.

Retrieval of variant sequences

The lineages of variants of concern, UK B.1.1.7 [38], South African B.1.351 [39], Brazilian P.1 [40] and Indian B.1.617.2 [41], were used to identify the most prevalent mutations for nsp3, nsp12, ORF3a and ORF9b that would be present in at least 75% of the sequences [42] (**Table 1**).

CD8+ T cell epitope prediction and immunogenicity modeling

Using the best tool-attribute combination, CD8+ T cell epitopes were predicted for nsp3, nsp12, ORF3a and ORF9b variants. Then, IEDB immunogenicity tool was used to predict immunogenicity scores for the top 300 duplicate-free epitopes from each variant [43]. Epitopes with an immunogenicity score >0 were selected for further analysis.

CD4+ T cell epitope prediction

CD4+ T cell variant epitopes were predicted for each respective protein using IEDB recommended 2.22 [44, 45]. Top 300 duplicate-free epitopes based on rank were selected for further analysis.

IFN- γ epitope prediction

IFN- γ plays a center role in viral infections for both innate and adaptive immunity. Therefore, it is important to identify IFN- γ inducing epitopes. IFNEpitope server was used to predict IFN- γ epitopes [46]. Only epitopes with an IFN- γ score >0 were selected.

Antigenicity, allergenicity and toxicity evaluation

Antigenicity was evaluated by VaxiJen 2.0 [47] server with a threshold of 0.4, allergenicity by AllerTop v2.0 [48] and toxicity by ToxinPred [49]. Peptides that were “probable non-allergenic” and non-toxic were selected.

Amino acid physiochemical properties

Evaluation of various physicochemical properties is essential for determining the safety and efficacy of the candidate vaccine. Hence, various chemical and physical parameters associated with the vaccine were predicted in this study using ProtParam tool on ExPASy [50]. Instability index, aliphatic index, GRAVY and half-life of top CD8 and CD4 epitopes were assessed. To make sure the predicted peptides are stable and not being degraded, instability index <40 was selected as threshold. The aliphatic index points to thermostability, higher the aliphatic index of a protein, greater is its thermostability. The Grand

average hydrophobicity (GRAVY) indicates the solubility of the peptide. Finally, the half-life in mammalian reticulocytes in vitro was predicted and half-life >1 hour was selected as a threshold value [51].

Worldwide human population coverage analysis

The population coverage of HLA class I and II alleles, calculated using the IEDB population coverage tool, computes the percentage of the world population predicted to present our epitopes on their MHC molecules. The analysis covered East Asia, Northeast Asia, South Asia, Southeast Asia, Southwest Asia, Europe, East Africa, West Africa, Central Africa, North Africa, South Africa, West Indies, North America, Central America, South America and Oceania [52].

Murine model MHC restriction

NetM2pan Server [53] was used to predict CD8 epitopes' affinity to murine MHC class I (H2) molecules. Two common mouse strains to study COVID-19 vaccines, C57BL/6 and BALB/CJ, were used in mouse HLA affinity predictions. Peptides that were either weak binders (WB) or strong binders (SB) were selected.

3D dimensional protein structure prediction

The tertiary structure of the top epitopes was predicted using PEPstrMOD [54]. The 3D structure of the respective HLA alleles which nsp3, nsp12, ORF3a and ORF9b epitopes bind to were downloaded from Protein Database Bank on RCSB.org. The native peptides from the RCSB structures were replaced with our predicted epitopes using PyMol. The epitopes were then docked into their respective MHC molecules using the FlexPepDock server [55].

Visualization of the top CD8 epitopes in 3D structure

The full-length reference protein sequences of ORF3a, nsp3, nsp12 and ORF9b proteins were retrieved from RCSB.org. The proteins and corresponding top CD8 epitopes were visualized using PyMol (**Supplementary Figures 1a-d**).

Results

Workflow of T cell predictions

A schematic presentation of the immunoinformatic methods was proposed for epitope predictions (**Figure 3**). The workflow of this study is illustrated in two main steps: IEDB tool comparison and MHC class I and

II epitope predictions. Briefly, the sequences of nsp3, nsp12, ORF3a and ORF9b were retrieved from SARS-CoV-2 to predict CD4 and CD8 epitopes. IEDB tool comparison was carried out to identify the best-performing tool-attribute combination for CD8 epitope predictions. Specifically, the nsp3 protein was selected for analysis due to its highest number of high-quality experimental epitopes. After identification of the best prediction tool, CD8 epitopes were screened through IEDB immunogenicity prediction tool, as well as antigenicity, allergenicity, toxicity, IFN- γ secretion and several physiochemical properties predictors to select the most effective vaccine constructs. Similar procedure was carried out for CD4 epitopes, however instead of carrying out tool comparisons, the best tool was selected from literature [56].

Tool comparisons for CD8+ T cell binding predictions

The performance of each attribute-tool combination was evaluated by generating receiver operating characteristic curves (ROC) and box plot graphs for each stringency level method (**Figure 4**). A ROC is a probability curve that utilizes Area Under the Curve (AUC) to visualize the performance of a tool-attribute combination [57]. Generally, the higher the AUC value, the better the model is performing. The area under the curve (AUC) values were compared for each stringency level and it was found that IEDB recommended (NetMHCpan EL 4.1 rank) had the highest combined AUC value, meaning its performance was most specific and sensitive (**Figure 4**). Therefore, MHC-I restricted T cell epitopes and their binding affinity were predicted using this tool.

Variant sequences

In order to design the most effective vaccine, it is important to identify the most prevalent mutations across the world. Since the start of the pandemic, four variants of concern (B.1.351, B.1.617.2, B.1.1.7, P.1) have been reported that carry mutations important to virus transmissibility [12]. After filtering through mutations that are presented by at least 75% of the sequences, we were able to identify several nonsynonymous mutations in the nsp3 protein (**Table 1**). On the other hand, the nsp12 protein that is involved in virus replication, was found to have a dominant mutation across all variants of concern (**Table 1**). The point mutation at P323L, first detected in the UK, has been associated with severity of COVID-19 [58]. Also, the amino acid mutation in Q57H of ORF3a has been correlated with severity of the disease [9]. In ORF9b protein, no prevalent mutations were found across the variants of concern.

CD8+ T cell epitope prediction

Using NetMHCpan EL 4.1 based on rank, MHC-I immunogenicity predictor, IFNEpitope, VaxiJen 2.0, AllerTop 2.0, ToxinPred and ProtParam, only immunogenic, antigenic, IFN- γ inducing, non-allergenic, non-toxic and stable CD8 epitopes were selected. We obtained 8 nsp3, 4 nsp12, 11 ORF3a and 3 ORF9b CD8 epitopes, common across all four variants, 1 nsp3 epitope specific to the UK variant and 4 ORF3a epitopes specific to the Indian variant (**Table 2**). While IFNEpitope server is mostly designed for MHC class II restricted epitopes, the MHC class I predictions were also made without the IFN- γ filter. We were able to obtain 21 common nsp3, 19 common nsp12, 36 common ORF3a and 5 common ORF9b CD8 epitopes across VOC, 2 nsp3 epitopes specific to the UK variant, 2 nsp12 epitopes specific to the Indian, the UK, Brazilian and South African variants, and 7 ORF3a epitopes specific to the Indian variant (**Supplementary Tables 2-6**).

CD4+ T cell epitope prediction

IEDB recommended 2.22 (rank) was selected as the best performing tool-attribute combination for variant CD4 epitope predictions due to its sensitivity and specificity computed in a previous study [56]. IFNEpitope, VaxiJen 2.0, AllerTop 2.0, ToxinPred and ProtParam, only antigenic, IFN- γ inducing, non-allergenic, non-toxic and stable CD4 epitopes were selected. We obtained 28 common nsp3, 19 common nsp12, 11 common ORF3a and 4 common ORF9b epitopes across VOC, 1 nsp3 epitope specific to the UK variant, 3 nsp3 epitopes specific to the Brazilian variant, 5 ORF3a epitopes specific to the South African variant and 3 ORF3a epitopes specific to the Indian variant (**Supplementary Table 7**).

Murine MHC restriction prediction

As murine models pose a great way to safely test the immunogenicity and efficacy of potential vaccine constructs, we identified top CD8 epitopes that have either strong or weak affinity to murine H2 molecules (**Table 3**). ORF3a was predicted to have 6 common immunogenic, antigenic, non-allergenic, non-toxic and stable CD8 epitopes across VOC with affinity to murine MHC restriction, while nsp3 and nsp12 had 5 and 4 CD8 epitopes, respectively (**Table 3**). Moreover, our analysis revealed three common immunogenic, antigenic, non-toxic, non-allergenic, stable and IFN- γ inducing CD8 epitopes presented by murine MHC restriction that overlap with at least one antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD4 epitope: GWLIVGVAL (ORF3a), TAFGLVAEW (nsp3) and VYANLGERV (nsp12). These CD8 epitopes were found to be within a total of 6 antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD4 epitopes: TAFGLVAEWFLAYIL (nsp3), LTAFGLVAEWFLAYI (nsp3), DLTAFGLVAEWFLAY (nsp3), PDILRVYANLGERVR (nsp12), RVYANLGERVRQALL (nsp12) and

PFGWLIVGVALLAVF (ORF3a) (**Table 4**). The world population coverage analysis revealed that GWLIVGVAL, TAFGLVAEW and VYANLGERV are able to elicit an immune response that covers 75.72% of the world population (**Supplementary Figure 2p**) while the combined 9 HLA-I and HLA-II epitopes are able to cover 99.99% of the world population (**Supplementary Figure 2q**). Therefore, we can state that the presented overlapping CD8 and CD4 epitopes could act as an excellent multi-epitope vaccine construct against SARS-CoV-2.

Population Coverage Analysis

CD8 epitopes

The population coverage of the T cell epitopes was estimated from the corresponding HLA associations. The population coverage was calculated for both datasets with and without IFNEpitope filter. The top common 8 nsp3, 4 nsp12, 11 ORF3a and 3 ORF9b epitopes across VOC cover 98.12%, 87.08%, 96.07% and 63.8% of the world population, respectively (**Table 2, Supplementary Figures 2e-h**), whereas all of them are immunogenic, antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing. Since IFNEpitope is more accurate for CD4 epitope predictions and in order to increase the population coverage, IFN- γ secretion filter was removed from the predictions. It was found that the 21 nsp3, 19 nsp12, 36 ORF3a and 5 ORF9b are able to elicit an immune response that covers 98.55% of world population in nsp3, nsp12 and ORF3a proteins, and 75.8% in ORF9b (**Supplementary Table 2, Supplementary Figures 2a-d**). Also, we computed the population coverage for our top common CD8 epitopes that are presented by murine model MHC restriction (**Table 3**). The fewer epitope-construct is still able to elicit a strong immune response that covers 97.45% of population in ORF3a, 95.11% in nsp3 and 91.62% in nsp12 world population (**Table 3, Supplementary Figures 2i-k**).

CD4 epitopes

The population coverage of CD4 epitopes was calculated for the few top epitopes. It was found that the top common 7 ORF3a, 3 nsp3, 6 nsp12 and 4 ORF9b epitopes are able to cover 99.76%, 99.82%, 96.62% and 94.74% of the world population, respectively (**Table 4, Supplementary Figures 2l-o**).

3D structure predictions

The top CD8 epitopes from each protein were selected and visualized with a respective HLA-I allele (**Figures 5-8**). TAFGLAEW was chosen as the best epitope in nsp3 due to its high antigenicity score of 1.0431, immunogenicity score of 0.19168, stability score of -7.81 (**Supplementary Table 3**) and its presence in two antigenic CD4 epitopes identified in this study (**Table 5**). For nsp12, SSVELKHFFF was

chosen due to its high antigenicity score of 1.4503 (**Supplementary Table 5**). GWLIVGVAL was selected as the best CD8 epitope in ORF3a due to its high immunogenicity score of 0.26012, positive IFN- γ score (**Supplementary Table 4**) and presence in an antigenic CD4 epitope (**Table 4**). Finally, a CD8 epitope DKAFQLTPI from ORF9b was chosen for 3D visualization due to its ability to interact with TOM70. Besides this, DKAFQLTPI was incorporated into two antigenic CD4 epitopes which suggests its higher potential to elicit a protective and strong immune response (**Table 4**). We also visualized the positions of selected top common immunogenic, antigenic, non-allergenic, non-toxic and stable CD8 epitopes across VOC in nsp3, nsp12 and ORF3a reference protein sequences (**Supplementary Figures 1a-c**). For nsp3, 5CD8 epitopes were visualized: LTAFLVLAEW, TAFGLVAEW, SYLFQHANL, EAFEYYHTT and LVSDIDITF (**Supplementary Figure 1b**). For nsp12, VYANLGERV, NPHLMGWDY, SSVELKHFFF and SAKNRARTV were visualized (**Supplementary Figure 1c**), and LLVAGLEA, GWLIVGVAL, GWLIVGVALL, NLLLLFVTV, GVEHVTFYIY, VEHVTFYIY, DTGVEHVTFY, GVEHVTFYI and TGVEHVTFY were visualized in ORF3a protein (**Supplementary Figure 1a**). The highest number CD8 epitopes from ORF3a protein was selected for visualization due to its highest number of common immunogenic CD8 epitopes.

Discussion

Built on the basis of immunoinformatics, peptide-based vaccine design has recently become a novel approach towards the identification and design of immunogenic peptides [60]. It is becoming increasingly apparent that T cells, one of the main aspects of the adaptive immune system, may play a significant role in providing protective immunity against COVID-19. Understanding the repertoire of SARS-CoV-2-specific epitopes that can trigger a robust T cell response in humans, is highly important for protective and effective vaccine design (**Figure 1**). With most of the vaccine design attempts focusing on the structural proteins, notably the spike protein, our work, aims to identify immunogenic epitopes from previously uncharacterized ORF proteins. ORF1ab polyprotein, located on the 5' end of the SARS-CoV-2 genome (**Figure 2**), has been reported to be immunogenic due to its high allele capacity, as well as for its large gene region, comprising of multiple non-structural proteins [7]. PLpro and RdRP, two non-structural proteins from ORF1ab, have essential functions to the virus as they play an important role in viral replication. In addition to these, ORF3a protein, that induces apoptosis in transfected and infected cells, has also been recognized to have high immunogenicity in activating SARS-CoV-2-specific T cells.

Choosing the appropriate epitopes that would be safe but also trigger an immune response against an invading pathogen, is a critical step in the development of a multi-epitope-based vaccine. To ensure the reliability of the various prediction tools offered online, tool comparison was the first necessary step towards the identification of the best performing tool-attribute combination to predict T cell epitopes. Our analysis which relied on experimental epitopes matched to the predicted epitopes, revealed that NetMHCpan EL 4.1 ordered based on rank was the best performing tool for CD8 epitope predictions, whereas IEDB recommended 2.22 ordered based on rank was found to perform best for CD4 epitope predictions.

In the present study, we have identified several immunogenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD8+ T cell epitopes and antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD4+ T cell epitopes from SARS-CoV-2 proteins. NSP3 which acts as papain-like protease protein, is responsible for the cleavage of several other non-structural proteins in SARS-CoV-2, an indication of its importance in the virus life cycle [61]. Presence of a macro domain in NSP3 which is involved in disruption of the expression of innate immunity genes, is another indication of the importance of the protein [62]. Our analysis revealed a high number of T cell epitopes in NSP3 with a total of 8 CD8 (**Supplementary Table 3**) and 28 CD4+ T cell epitopes, common to the VOC (**Supplementary Table 7**). The predicted CD8 epitopes had a high presentation of HLA alleles which may contribute to induced immunogenicity [63]. Moreover, we identified 4 antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD4 epitopes that had a corresponding immunogenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD8+ T cell epitope within it (**Table 5**). The combination of CD4 and CD8 T cell epitopes indicates a potential more powerful immune response against SARS-CoV-2.

ORF3a protein, known for the suppression of innate immune response [64], was also predicted to have a high number of T cell epitopes for an effective vaccine design. We predicted a total of 11 immunogenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD8 and 11 antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD4 epitopes in ORF3a protein, common across VOC (**Supplementary Tables 4,7**). ORF3a top common epitopes were also predicted to have a wide range of HLA-I and HLA-II binding capacity. Out of the 11 CD4+ T cell epitopes, 7 of them had an immunogenic CD8+ T cell within its sequence (**Table 5**) which indicates these epitopes from ORF3a have a high potential of triggering a strong protective immune response against SARS-CoV-2. Additionally, the population coverage of the 7 IFN- γ inducing CD4+ T epitopes was predicted to be 99.76% of the world population (**Table 4**), making ORF3a a desirable vaccine candidate. The results suggested that NSP12 had the lowest number of immunogenic CD8+ T cell epitopes with the lowest population coverage, mainly due to its shorter sequence (932 aa) compared to NSP3 (**Table 2**). However, as CD8 epitope predictions were conducted

without the IFN- γ server, we were able to obtain 15 additional common immunogenic, antigenic, non-allergenic, non-toxic and stable epitopes (**Supplementary Table 5**) which subsequently boosted the population coverage to 98.55% (**Supplementary Table 2**).

ORF9b is an important source of T cell epitopes that remains largely unexplored in the context of T-cell immunity. Although relatively short (97 aa), ORF9b yielded three immunogenic, non-allergenic, non-toxic, stable and IFN- γ inducing HLA-I peptides: HPALRLVDP, DKAFQLTPI and AVTRMENA that cover 63.8% of the world population (**Supplementary Table 6**). To make the outcome more specific, we aimed to identify immunogenic epitopes from ORF9b that could bind to TOM70. Previous studies report that ORF9b occupies the hydrophobic pocket on the TOM70 C-terminal domain (CTD), suppressing the IFN-I production. ORF9b is also expressed early in infection which points to better HLA-I presentation and immunogenicity [10]. In the study, we were able to identify one ORF9b CD8+ T cell epitope (DKAFQLTPI) in the range of the interaction site (aa 43-78) with TOM70 (PDB:7DHG; **Supplementary Figure 1d**). DKAFQLTPI that binds only to HLA-B*51:01, was shown to be immunogenic, highly antigenic, non-allergenic, non-toxic, IFN- γ inducing and stable. Besides this, we identified 3 antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD4+ T cell epitopes – LEDKAFQLTPIAVQM, SLEDKAFQLTPIAVQ and NSLEDKAFQLTPIAV that had DKAFQLTPI incorporated into the sequence (**Table 5**). Although, DKAFQLTPI has no weak or strong binding affinity to murine MHC restriction, it could be used as a part of the multi-epitope vaccine. The high quality CD8+ T cell epitope with booster CD4 epitopes could provide a strong immune response early in infection and inhibit the virus from evading. Considering the critical role of IFN-I in the human antiviral response, re-creation of IFN-I production in COVID-19 infected individuals may prove to be a significantly effective therapeutic option. Since ORF9b is also highly homologous to ORF9b protein in SARS-CoV, the vaccine construct could be used against other SARS-like coronaviruses [10].

A limitation of current studies relates to mutational spectra of SARS-CoV-2 that is constantly changing. This poses a risk to the vaccine candidates to no longer being viable to offer protection against newly formed variants. To overcome this, we selected the most prevalent variants of concern and their most common mutations in nsp3, nsp12, ORF3a and ORF9b proteins across all identified sequences to date. The selected proteins may carry some important mutations that increase the transmissibility and survival of the virus, like the nonsynonymous mutation P323L in nsp12 which has possibly affected the catalytic activity of the nsp12 protein, therefore altering the binding capability with other cofactors, such as nsp7 and nsp8 proteins [65]. However, studies have reported generally low mutation rates in these proteins which increase the efficiency of the proposed multi-epitope vaccine candidates in the long term [66]. Another challenge with peptide vaccines is that peptides can be easily degraded by peptidases which

causes a lower T cell response. The present study addresses this issue by calculating the half-life of the top common epitopes to make sure that the vaccine construct is not degraded early in blood.

Conclusion

With the lack of effective antiviral treatments, there is an urgent need for a protective vaccine that is effective against different strains of SARS-CoV-2 and among different ethnic groups across the world. Our work sheds light on previously uncharacterized SARS-CoV-2 HLA-I and HLA-II peptides from ORF proteins in the SARS-CoV-2 genome which reveals multiple T cell epitopes that could be used in peptide-based vaccine construction. To ensure the vaccine construct is effective against most common circulating mutations, most prevalent mutations from four lineages of variants of concern: B.1.1.7, P.1, B.1.617.2 and B.1.351 were collected and analyzed. We predicted CD4 and CD8 T cell epitopes for two non-structural proteins, nsp3 and nsp12, as well as from ORF3a and ORF9b proteins. We identified 5 immunogenic, antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing nsp3 CD8 epitopes with at least weak affinity to one or more mouse MHC alleles, 4 for nsp12 protein and 6 for ORF3a protein, all common to the studied VOC. Besides this, the variant specific epitopes could be used in variant-specific booster vaccine constructs. Our analysis also revealed 3 immunogenic, antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD8 epitopes with affinity to mouse MHC alleles and presence within at least one antigenic non-allergenic, non-toxic, stable and IFN- γ inducing CD4 epitope which altogether have the ability to provoke a strong immune response and protect 99.99% of the world population, suggesting its efficiency as potential multi-epitope vaccine construct. ORF9b, the open-reading frame protein within nucleocapsid protein, was predicted to have one immunogenic, antigenic, non-allergenic, non-toxic, stable and IFN- γ inducing CD8 epitope, DKAFQLTPI, that could suppress the interaction between ORF9b and human mitochondrial outer membrane protein, TOM70. These findings suggest that a single multi-epitope vaccine candidate should be efficacious against currently circulating lineages.

Acknowledgements

We wish to acknowledge for the support by Lombardi Comprehensive Cancer Center, Georgetown University Medical Center.

References

1. Noorimotlagh Z, Karami C, Mirzaee SA, Kaffashian M, Mami S, Azizi M. Immune and bioinformatics identification of T cell and B cell epitopes in the protein structure of SARS-CoV-2: A systematic review. *Int Immunopharmacol*. 2020;86:106738. Accessed August 27, 2021. doi: 10.1016/j.intimp.2020.106738.
2. Ugurel OM, Ata O, Turgut-Balik D. An updated analysis of variations in SARS-CoV-2 genome. *Turk J Biol*. 2020;44(3):157-167. Accessed Sept 5, 2021. doi:10.3906/biy-2005-111.
3. WHO Coronavirus (COVID-19) Dashboard. *World Health Organization*. Accessed Sept 3, 2021. <https://covid19.who.int/>
4. Gershoni JM, Roitburd-Berman A, Siman-Tov DD, et al. Epitope Mapping. *BioDrugs* 2007;21, 145-156. Accessed September 14, 2021. doi: 10.2165/00063030-200721030-00002.
5. Laidlaw B, Craft J, Kaech S. The multifaceted role of CD4⁺ T cells in CD8⁺ T cell memory. *Nat Rev Immunol* 2016;16, 102–111. Accessed September 5, 2021. doi: 10.1038/nri.2015.10.
6. Khan MI, Khan ZA, Baig MH, et al. Comparative genome analysis of novel coronavirus (SARS-CoV-2) from different geographical locations and the effect of mutations on major target proteins: An *in silico* insight. *PLoS ONE* 15 2020;(9):e0238344. Accessed September 4, 2021. doi: 10.1371/journal.pone.0238344.
7. Tarke A, Sidney J, Kidd CK, et al. Comprehensive analysis of tcell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep Med*. 2021;2(2):100204. Accessed June 13, 2021. doi: 10.1016/j.xcrm.2021.100204.
8. Hisham Y, Ashhab Y, Hwang SH, Kim DE. Identification of Highly Conserved SARS-CoV-2 Antigenic Epitopes with Wide Coverage Using Reverse Vaccinology Approach. *Viruses*. 2021;13(5):787. Accessed June 14, 2021. doi:10.3390/v13050787.
9. Majumdar P, Niyogi S. ORF3a mutation associated with higher mortality rate in SARS-CoV-2 infection. *Epidemiol Infect*. 2020;148:e262. Accessed September 4, 2021. doi:10.1017/S0950268820002599.
10. Jiang Hw, Zhang Hn, Meng Qf. et al. SARS-CoV-2 Orf9b suppresses type I interferon responses by targeting TOM70. *Cell Mol Immunol* 2020;17, 998–1000 Accessed August 12, 2021. doi: 10.1038/s41423-020-0514-8.

11. Han L, Zhuang M, Zheng Y, et al. SARS-CoV-2 ORF9b Antagonizes Type I and III Interferons by Targeting Multiple Components of RIG-I/MDA-5-MAVS, TLR3-TRIF, and cGAS-STING Signaling Pathways. *bioRxiv*; 2020. Accessed September 8, 2021. doi: 10.1101/2020.08.16.252973.
12. Khateeb J, Li Y, Zhang H. Emerging sars-cov-2 variants of concern and potential intervention approaches. *Crit Care* 2021;25, 244. Accessed September 4, 2021. doi: 10.1186/s13054-021-03662-x.
13. Peters B, Sidney J, Bourne P, et al. The design and implementation of the immune epitope database and analysis resource. *Immunogenetics* 2005;57, 326–336. Accessed June 14, 2021. doi: 10.1007/s00251-005-0803-5.
14. Johansen MD, Irving A, Montagutelli X. et al. Animal and translational models of SARS-CoV-2 infection and COVID-19. *Mucosal Immunol* 2020;13, 877-891. Accessed September 8, 2021. doi: 10.1038/s41385-020-00340-z.
15. Zhang Y, Huang K, Wang T, et al. SARS-CoV-2 rapidly adapts in aged BALB/c mice and induces typical pneumonia. *J Virol*. 2021;95(11):e02477-20. Accessed August 27, 2021. doi:10.1128/JVI.02477-20.
16. Ferretti P, Kula TY, Wang SA, et al. Unbiased screens show CD8+ T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that largely reside outside the spike protein. *Immunity* 2020;53(5):1095-1107.e3. Accessed June 14, 2021. doi: 10.1016/j.immuni.2020.10.006.
17. Grifoni A, Weiskopf D, Ramirez SI, et al. Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell*. 2020; 181(7):1489-1501. Accessed June 14, 2021. doi: 10.1016/j.cell.2020.05.015.
18. Kared H, Redd AD, Bloch EM, et al. CD8+ T cell responses in convalescent COVID-19 individuals target epitopes from the entire SARS-CoV-2 proteome and show kinetics of early differentiation. Preprint. *bioRxiv*. 2020;2020.10.08.330688. Published Oct 8, 2020. Accessed June 15, 2021. doi:10.1101/2020.10.08.330688.
19. Saini SK, Hersby DS, Tamhane T, et al. SARS-CoV-2 genome-wide mapping of CD8 T cell recognition reveals strong immunodominance and substantial CD8 T cell activation in COVID-19 patients. *Science immunology*. 2021. Accessed June 14, 2021. doi: 10.1126/sciimmunol.abf7550 2021.

20. Schulien J, Kemming V, Oberhardt K, et al. Characterization of pre-existing and induced SARS-CoV-2-specific CD8+ T cells. *Nat. Med.* 2020;27, 78–85. Published November 20, 2020. Accessed June 14, 2021. doi: 10.1038/s41591-020-01143-2.
21. Snyder TM, Gittelman RM, Klinger M, et al. Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. Preprint. *medRxiv.* 2020;2020.07.31.20165647. Published Sep 17, 2020. Accessed June 14, 2021. doi:10.1101/2020.07.31.20165647.
22. Takagi A, Matsui M. Identification of HLA-A*02:01-restricted candidate epitopes derived from the non-structural polyprotein 1a of SARS-CoV-2 that may be natural targets of CD8+ T cell recognition *in vivo*. *J Virol.* 2020;95(5):e01837-20. Accessed August 6, 2021. doi:10.1128/JVI.01837-20.
23. Tarke A, Sidney J, Kidd CK, et al. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *bioRxiv.* 2020. Accessed June 12, 2021. doi: 10.1101/2020.12.08.416750.
24. Kumar M, et. al. Experimental known SARS-CoV T-cell epitopes mapped to SARS-CoV-2 (2019-nCoV). 2020. Accessed June 20, 2021. <https://bioinfo.imtech.res.in/manojk/coronavr/tcellSARS-ncov.php>
25. Nielsen M, Lundegaard C, Blicher T, et al. NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS One.* 2007;2(8):e796. Accessed June 13, 2021. doi: 10.1371/journal.pone.0000796.
26. Hoof I, Peters B, Sidney J, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics.* 2009;61(1):1-13. Accessed June 13, 2021. doi: 10.1007/s00251-008-0341-z.
27. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* 2016;8(1):33. Accessed June 13, 2021. doi: 10.1186/s13073-016-0288-x.
28. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved peptide-MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol.* 2017;199(9):3360-3368. Accessed June 13, 2021. doi: 10.4049/jimmunol.1700893.

29. Reynisson B, Alvarez B, Paul S, Peters B, Nielsen M. NetMHCpan-4.1 and NetMHCIIpan-4.0: Improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 2020;48(W1):W449-W454. Accessed June 13, 2021. doi: 10.1093/nar/gkaa379.
30. Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics.* 2016;32(4):511-517. Accessed June 13, 2021. doi: 10.1093/bioinformatics/btv639.
31. Buus S, Lauemøller SL, Worning P, et al. Sensitive quantitative predictions of peptide-MHC binding by a 'query by committee' artificial neural network approach. *Tissue Antigens.* 2003;62(5):378-384. Accessed June 13, 2021. doi: 10.1034/j.1399-0039.2003.00112.x.
32. Lundegaard C, Nielsen M, Lund O. The validity of predicted T-cell epitopes. *Trends Biotechnol.* 2006;24(12):537-538. Accessed June 13, 2021. doi: 10.1016/j.tibtech.2006.10.001.
33. Lundegaard C, Lund O, Nielsen M. Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics.* 2008;24(11):1397-1398. Accessed June 13, 2021. doi: 10.1093/bioinformatics/btn128.
34. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: Accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-35. *Nucleic Acids Res.* 2008;36(Web Server issue):509. Accessed June 13, 2021. doi: 10.1093/nar/gkn202.
36. Nielsen M, Lundegaard C, Worning P, et al. Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* 2003;12(5):1007-1017. Accessed June 13, 2021. doi: 10.1110/ps.0239403.
37. Moutaftsi M, Peters B, Pasquetto V, et al. A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nat Biotechnol.* 2006;24(7):817-819. Accessed June 13, 2021. doi: 10.1038/nbt1215.
38. Temsah MH, Barry M, Aljamaan F, et al. SARS-CoV-2 B.1.1.7 UK Variant of Concern Lineage-Related Perceptions, COVID-19 Vaccine Acceptance and Travel Worry Among Healthcare Workers. *Front Public Health.* 2021;9:686958. Published 2021 May 26. Accessed September 10, 2021. doi:10.3389/fpubh.2021.686958.

39. Mwenda M, Saasa N, Sinyange N, et al. Detection of B.1.351 SARS-CoV-2 Variant Strain - Zambia, December 2020. *MMWR Morb Mortal Wkly Rep.* 2021;70(8):280-282. Published 2021 Feb 26. Accessed September 10, 2021. doi:10.15585/mmwr.mm7008e2.
40. Silva MSD, Demoliner M, Hansen AW, et al. Early detection of SARS-CoV-2 P.1 variant in Southern Brazil and reinfection of the same patient by P.2. *Rev Inst Med Trop Sao Paulo.* 2021;63:e58. Published 2021 Jul 5. Accessed September 10, 2021. doi:10.1590/S1678-9946202163058.
41. Planas D, Veyer D, Baidaliuk A, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* 596, 276–280 (2021). Accessed September 10, 2021. doi: 10.1038/s41586-021-03777-9.
42. Mullen J, Tsueng G, et int., and tCfVSB. “Outbreak.info.” Published in 2020. Accessed June 17, 2021. *outbreak.info*
43. Calis JJA, Maybeno M, Greenbaum JA, et al. Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS Computational Biology*, 2013;vol. 9, no. 10. Accessed September 3, 2021. doi: [10.1371/journal.pcbi.1003266](https://doi.org/10.1371/journal.pcbi.1003266).
44. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics.* 2020;11:568. Accessed Jun 17, 2021. doi: 10.1186/1471-2105-11-568.
45. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B. 2008. A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Comput Biol.*4(4):e1000048. Accessed Jun 17, 2021. doi: 10.1371/journal.pcbi.1000048.
46. Dhanda S. K, Vir P, Raghava GPS. Designing of Interferon-Gamma Inducing MHC Class-II Binders. *Biol. Direct.* 2013;8:30. Accessed June 19, 2021. doi: 10.1186/1745-6150-8-30.
47. Doytchinova IA, Flower DR. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics.* 2007;8:4. Accessed June 19, 2021. doi: 10.1186/1471-2105-8-4.
48. Dimitrov I, Flower DR. and Doytchinova I. AllerTOP - a server for *in silico* prediction of allergens. *BMC Bioinformatics* 14, S4 (2013). Accessed June 19, 2021. doi: 10.1186/1471-2105-14-S6-S4.

49. Gupta S, Kapoor P, Chaudhary K, et al. In Silico Approach for Predicting Toxicity of Peptides and Proteins. *PLoS ONE*. 2013;8:e73957. Accessed June 19, 2021. doi: 10.1371/journal.pone.0073957.
50. Gasteiger E, Hoogland C, Gattiker A, et al. Protein identification and analysis tools on the ExPASy server. In: John M. Walker, ed. *The proteomics protocols handbook*. Totowa, N.J: Humana Press; 2005:571-607. Accessed June 19, 2021.
51. Dey AK, Malyala P, Singh M. Physicochemical and functional characterization of vaccine antigens and adjuvants. *Expert Rev Vaccines*. 2014;13(5):671-685. Accessed June 21, 2021. doi:10.1586/14760584.2014.907528.
52. Bui HH, Sidney J, Dinh K, Southwood S, Newman MJ, Sette A. Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinformatics* 2006;17:153. Accessed September 2, 2021. doi: 10.1186/1471-2105-7-153.
53. DeVette CI, Andreatta M, Bardet W, et al. NetM2pan: A computational tool to guide MHC peptide prediction on murine tumors. *Cancer Immunol Res*. 2018;6(6):636-644. Accessed August 12, 2021. doi: 10.1158/2326-6066.CIR-17-0298.
54. Kaur H, Garg A, Raghava GPS. PEPstr: A de novo method for tertiary structure prediction of small bioactive peptides. *Protein Pept Lett*. 2007;14(7):626-631. Accessed August 21, 2021. doi: 10.2174/092986607781483859.
55. London N, Raveh B, Cohen E, Fathi G, Schueler-Furman O. Rosetta FlexPepDock web server-- high resolution modeling of peptide-protein interactions. *Nucleic Acids Res*. 2011;39(Web Server issue):249. Accessed August 15, 2021. doi: 10.1093/nar/gkr431.
56. Jabbour G, Rego S, Nguyenkhoa V, Dakshanamurthy S. Design of T cell epitope-based vaccine candidate for SARS-CoV-2 targeting spike and nucleocapsid protein escape variants. bioRxiv 2021.09.11.459907; Accessed September 13, 2021. doi: 10.1101/2021.09.11.459907.
57. Narkhede S. Understanding AUC - ROC CURVE. Published June 26, 2018. Accessed September 3, 2021. <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5>
58. Biswas SK, Mudi SR. Spike protein D614G and RdRp P323L: the SARS-CoV-2 mutations associated with severity of COVID-19. *Genomics Inform*. 2020;18(4):e44. Accessed August 13, 2021. doi: 10.5808/GI.2020.

59. Adapted from “Genome Organization of SARS-CoV”, by *BioRender.com* (2021). Retrieved from <https://app.biorender.com/biorender-templates>
60. Lim HX, Lim J, Poh CL. Identification and selection of immunodominant B and T cell epitopes for dengue multi-epitope-based vaccine. *Med Microbiol Immunol*. 2021 Feb;210(1):1-11. doi: 10.1007/s00430-021-00700-x. Epub 2021 Jan 30. PMID: 33515283.
61. Lei J, Kusov, Y, Hilgenfeld R. Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antiviral Res*. 2018;149, 58–74. Accessed September 5, 2021. doi: 10.1016/j.antiviral.2017.11.001
62. Fehr AR, Channappanavar R, Jankevicius G, Fett C, Zhao J, et al. The conserved coronavirus macrodomain promotes virulence and suppresses the innate immune response during severe acute respiratory syndrome coronavirus infection. *mBio* 2016;7:13. Accessed September 5, 2021. doi: 10.1128/mBio.01721-16
63. Sette A, Vitiello A, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J. Immunol*. 1994;153: 5586–5592. Accessed September 9, 2021. <http://www.jimmunol.org/content/153/12/5586>
64. Redondo N, Zaldivar-Lopez S, Garrido JJ, Montoya M. SARS-CoV-2 Accessory Proteins in Viral Pathogenesis: Knowns and Unknowns. *Front. Immunol*. 2021;12:708264. Accessed September 5, 2021. doi: 10.3389/fimmu.2021.708264.
65. Pachetti M, Marini B, Benedetti F, et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 2020;18, 179. Accessed September 10, 2021 doi: 10.1186/s12967-020-02344-6
66. Vilar S, Isom DG. One Year of SARS-CoV-2: How Much Has the Virus Changed? *Biology* 2021,10,91. Accessed September 10, 2021. doi: 10.3390/biology10020091

Figures and Tables

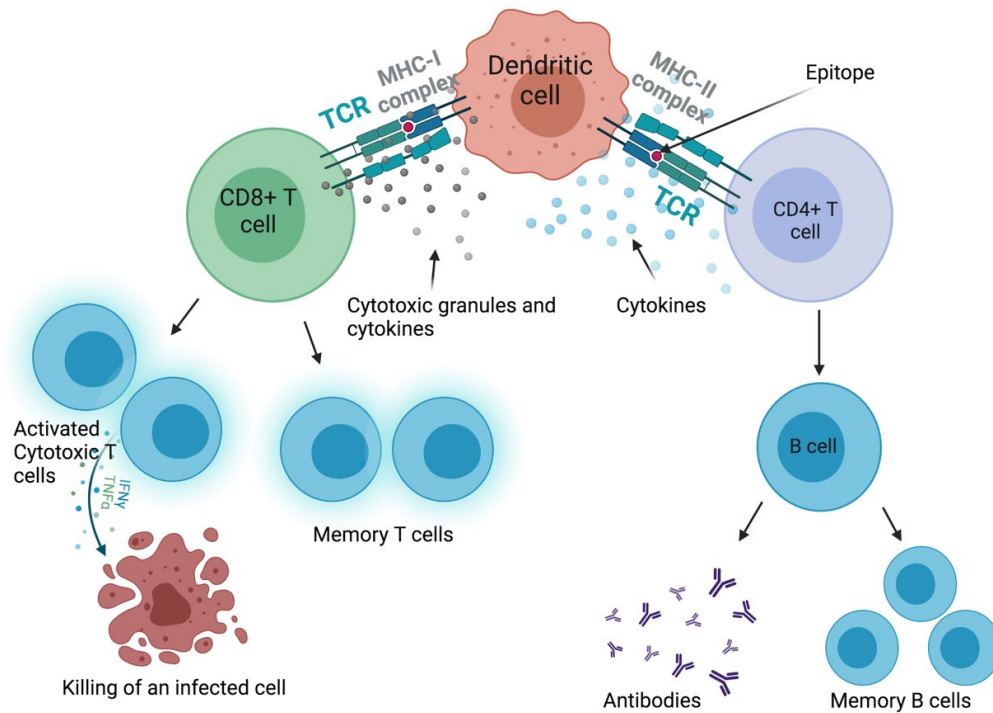


Figure 1. CD4+ and CD8+ T cell activation and response. Interaction between MHC molecules and T cell receptors (TCR) on T cells trigger the activation of CD4+ and CD8+ T cells that lead to the production of memory T and B cells. Cytokines and cytotoxic granules are released in response to a stimulus. The figure was created with BioRender.com.

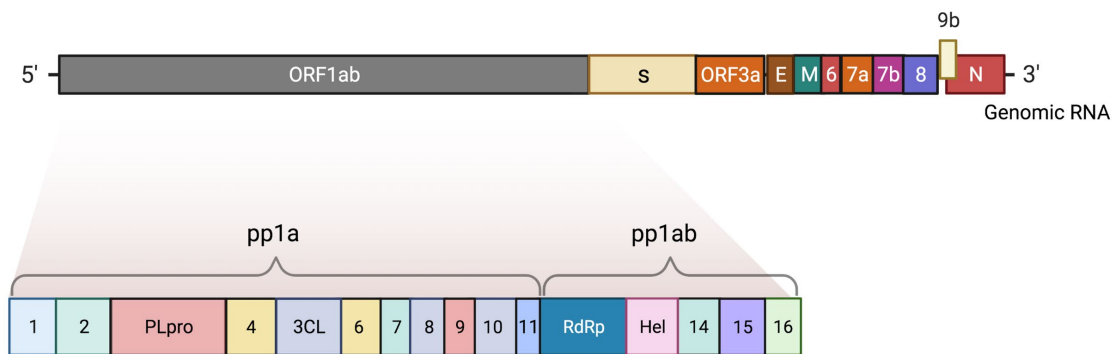


Figure 2. The complete genome of SARS-CoV-2. The 5' end consists of a large gene region ORF1ab and its non-structural proteins (nsp1-16). The 3' end comprises the structural proteins (nucleocapsid, membrane, envelope, spike) and other open-reading-frame (ORF) proteins. The figure was created with BioRender.com.

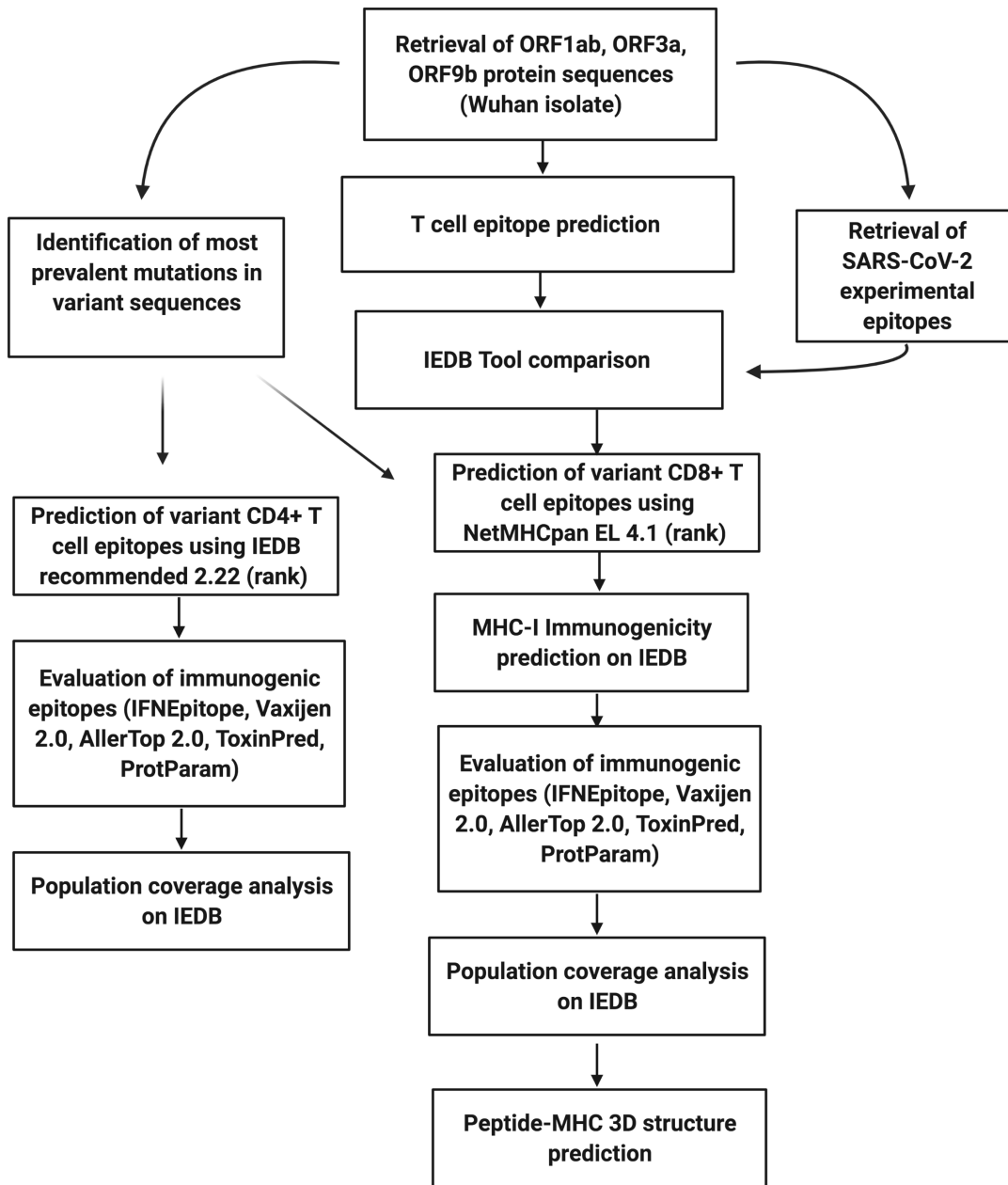


Figure 3. Schematic representation of immunoinformatics identification of T-cell epitopes in SARS-CoV-2. First, protein sequences were retrieved from SARS-CoV-2, which was followed by T cell epitope prediction and IEDB tool comparisons. Variant CD8 and CD4 epitopes were then predicted and analyzed based on their clinical properties. Finally, top common epitopes were used for population coverage and 3D prediction analysis. The figure was created with BioRender.com.

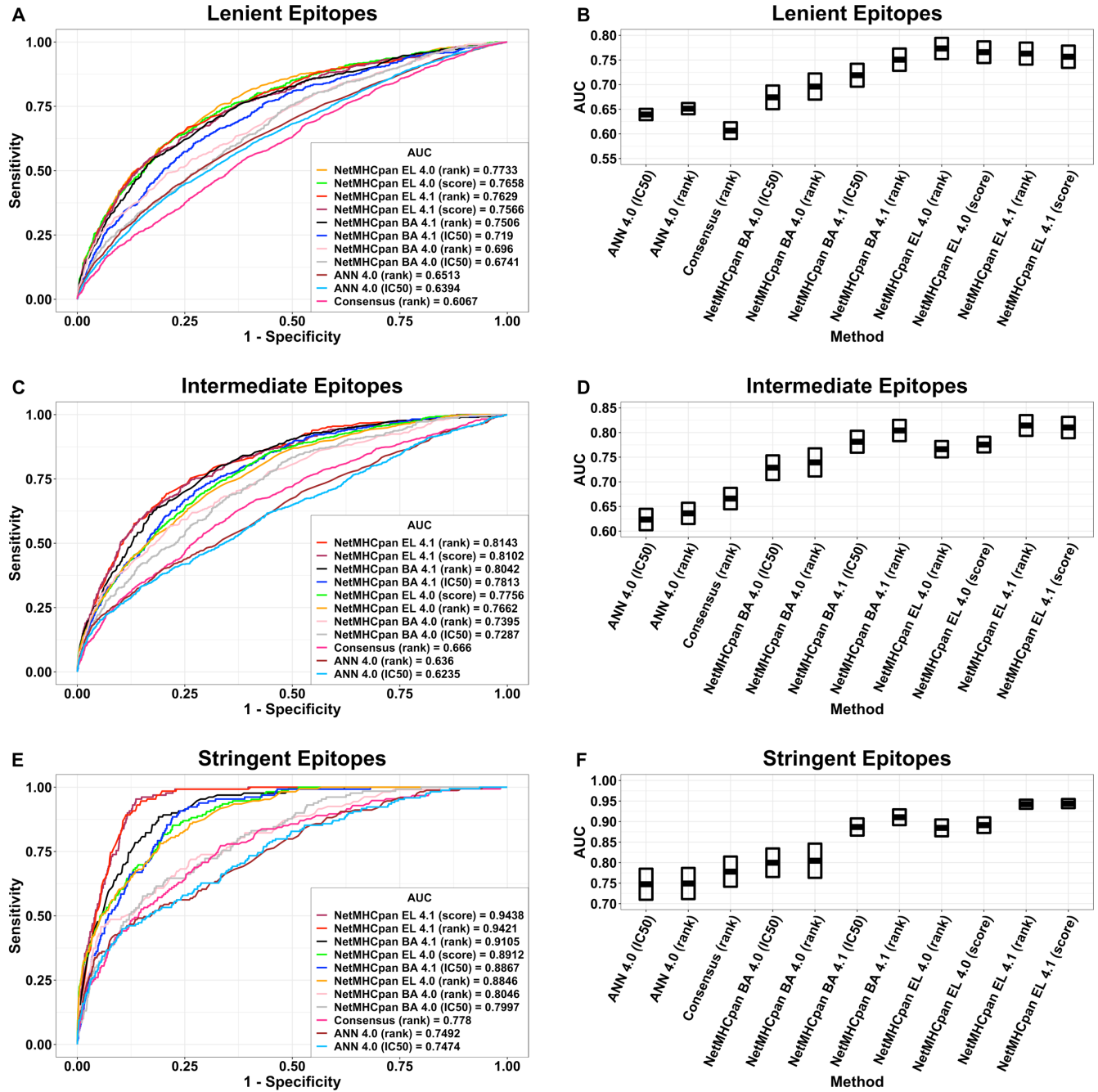


Figure 4. IEDB CD8+ T cell epitope tool comparisons using three levels of stringency: lenient, intermediate and stringent. ROC curves and box plot graphs illustrate the tool-attribute combinations that are useful in the selection of the best tool for CD8 epitope predictions.

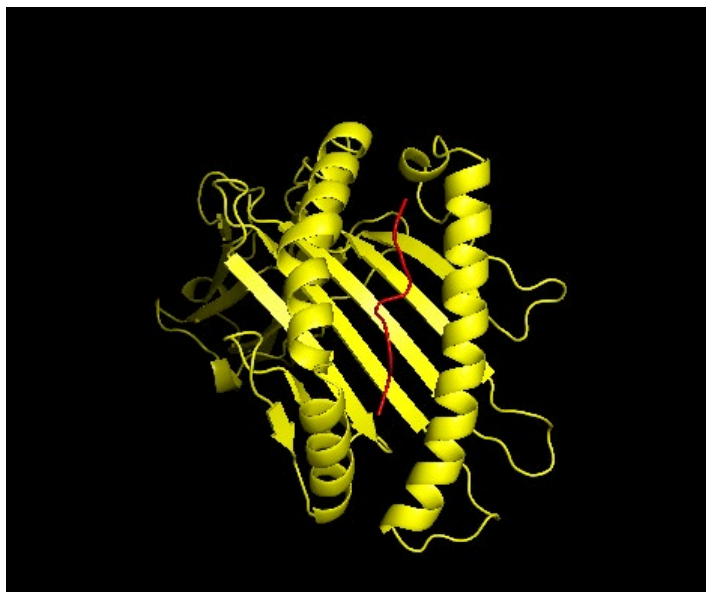


Figure 5. 3D structure of TAFGLAEW (red) in nsp3 protein and its interaction with HLA-B*58:01 allele (yellow)

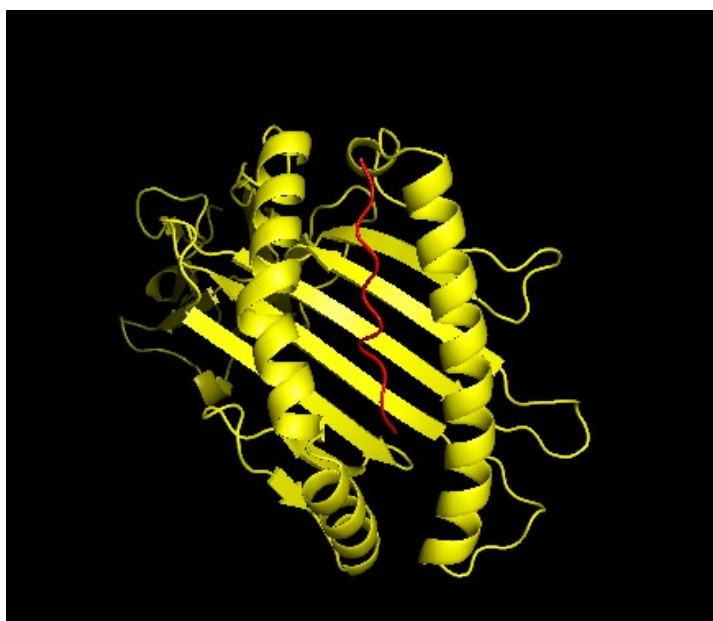


Figure 6. 3D structure of SSVELKHFFF (red) in nsp12 protein and its interaction with HLA-B*57:01 allele (yellow)



Figure 7. 3D structure of GWLIVGVAL (red) in ORF3a protein and its interaction with HLA-A*24:02 allele (yellow)

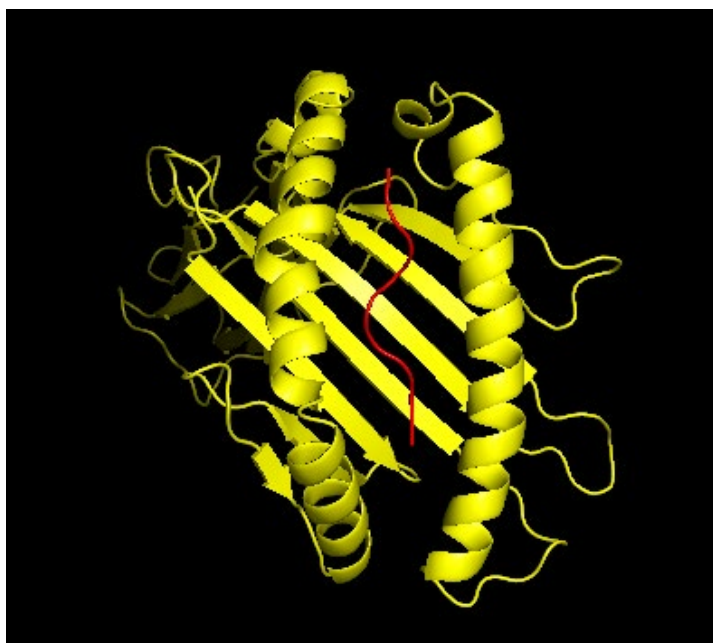


Figure 8. 3D structure of DKAFQLTPI (red) in ORF9b protein and its interaction with HLA-B*51:01 allele (yellow)

Variant	ORF1a (nsp3)	ORF1ab (nsp12)	ORF3a
South Africa (B.1.351)	K837N	P323L	Q57H, S171L
India (B.1.617.2)	A488S, P1228L, P1469S	P323L, G671S	S26L
UK (B.1.1.7)	T183I, A890D, I1412T	P323L	N/A
Brazil (P.1)	S370L, K977Q	P323L	S253P

Table 1. Most prevalent mutations in four variants of concern (VOC): B.1.351, B.1.617.2, B.1.1.7 and P.1. The point mutations are present in at least 75% of identified sequences.

Protein	Population Coverage	Epitope	HLA-I restriction
NSP3	98.12%	LTAFLGLVAEW	HLA-B*57:01, HLA-B*58:01, HLA-B*53:01, HLA-A*32:01, HLA-A*26:01, HLA-B*35:01
		LTAVVIPTK	HLA-A*11:01, HLA-A*68:01, HLA-A*03:01, HLA-A*30:01, HLA-A*31:01, HLA-A*33:01, HLA-A*26:01, HLA-B*57:01, HLA-B*58:01, HLA-A*32:01, HLA-A*01:01, HLA-A*30:02
		TAFGLVAEW	HLA-B*58:01, HLA-B*53:01, HLA-B*57:01, HLA-A*32:01, HLA-B*35:01, HLA-A*26:01, HLA-B*51:01, HLA-B*44:03, HLA-B*44:02, HLA-A*23:01, HLA-A*24:02, HLA-A*68:02, HLA-B*15:01, HLA-A*30:02, HLA-A*68:01, HLA-A*01:01, HLA-A*33:01
		EAFEYYHTT	HLA-A*68:02, HLA-B*51:01, HLA-A*26:01, HLA-B*35:01, HLA-B*08:01, HLA-B*53:01, HLA-A*33:01, HLA-A*02:06

		VVVNAANVY	HLA-A*30:02, HLA-A*26:01, HLA-B*35:01, HLA-A*01:01, HLA-B*53:01, HLA-A*32:01, HLA-A*11:01, HLA-B*58:01, HLA-B*57:01, HLA-A*03:01, HLA-A*68:01
		LEMELTPVV	HLA-B*40:01, HLA-B*44:03, HLA-B*44:02, HLA-A*02:06, HLA-B*51:01, HLA-A*02:01, HLA-B*53:01, HLA-B*35:01, HLA-B*08:01
		TLKGVEAVMY	HLA-B*15:01, HLA-A*30:02, HLA-A*26:01, HLA-A*01:01, HLA-A*03:01, HLA-B*35:01, HLA-A*32:01
		SYLFQHANL	HLA-A*24:02, HLA-A*23:01, HLA-B*08:01, HLA-A*32:01, HLA-A*30:01, HLA-A*33:01, HLA-A*31:01, HLA-A*30:02
NSP12	87.08%	VYANLGERV	HLA-A*24:02, HLA-A*23:01
		SAKNRARTV	HLA-B*08:01, HLA-B*51:01, HLA-A*30:01, HLA-B*07:02, HLA-A*02:03, HLA-A*30:02, HLA-A*26:01
		NPHLMGWDY	HLA-B*35:01, HLA-B*53:01, HLA-A*26:01, HLA-A*01:01, HLA-B*07:02, HLA-B*51:01
		MASLVLARK	HLA-A*68:01, HLA-A*11:01, HLA-A*03:01, HLA-A*30:01, HLA-A*33:01, HLA-A*31:01, HLA-B*58:01, HLA-A*26:01
ORF3a	96.07%	GWLIVGVALL	HLA-A*02:01, HLA-A*02:03, HLA-A*23:01, HLA-A*24:02
		GWLIVGVAL	HLA-A*23:01, HLA-A*24:02
		RATATIPIQA	HLA-A*30:01, HLA-A*68:02, HLA-A*02:06, HLA-B*58:01, HLA-A*30:02, HLA-A*02:03
		ATATIPIQA	HLA-A*68:02, HLA-A*02:06, HLA-A*30:01, HLA-A*02:03, HLA-A*02:01, HLA-A*30:02, HLA-A*11:01, HLA-A*01:01, HLA-B*58:01, HLA-A*03:01, HLA-A*68:01, HLA-A*32:01, HLA-A*26:01, HLA-B*57:01, HLA-B*40:01
		DFVRATATI	HLA-A*24:02, HLA-A*23:01, HLA-B*51:01, HLA-B*08:01, HLA-A*68:02
		LLLVAAGLEA	HLA-A*02:01, HLA-A*02:03
		LLVAAGLEA	HLA-A*02:03, HLA-A*02:01, HLA-A*02:06
		NLLLLFVTV	HLA-A*02:01, HLA-A*02:06, HLA-A*02:03, HLA-B*08:01, HLA-A*68:02, HLA-A*32:01, HLA-B*51:01
		LIVGVALLAV	HLA-A*02:03, HLA-A*02:06, HLA-A*02:01
		HLLLVAAGL	HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-B*08:01, HLA-A*32:01

		IVGVALLAV	HLA-A*02:06, HLA-A*02:01, HLA-A*02:03, HLA-A*68:02, HLA-A*32:01, HLA-B*51:01
ORF9b	63.8%	HPALRLVDP	HLA-B*07:02, HLA-B*08:01, HLA-B*35:01, HLA-B*51:01
		DKAFQLTPI	HLA-B*51:01
		AVTRMENA	HLA-A*02:06, HLA-A*68:02, HLA-A*02:03, HLA-A*02:01

Table 2. Top common IFN- γ inducing CD8+ T cell epitopes from nsp3, nsp12, ORF3a and ORF9b proteins across VOC and their population coverage. The selected epitopes are immunogenic, antigenic, non-allergenic, non-toxic and stable.

Protein	Population Coverage	Epitope	Mouse MHC restriction	Human HLA-I restriction
ORF3a	97.45%	DTGVEHVTFE	H-2-Dd	HLA-A*26:01, HLA-A*68:02, HLA-B*53:01, HLA-A*01:01, HLA-B*35:01, HLA-B*51:01, HLA-B*57:01, HLA-B*58:01
		TGVEHVTFE	H-2-Dd, H-2-Db, H-2-Kb	HLA-B*53:01, HLA-A*26:01, HLA-B*35:01, HLA-B*51:01, HLA-B*58:01, HLA-B*15:01, HLA-A*24:02, HLA-A*23:01, HLA-B*57:01, HLA-B*08:01, HLA-A*32:01, HLA-A*30:02, HLA-A*02:06

		INFVRIIMRL	H-2-Dd, H-2-Kb	HLA-A*23:01, HLA-A*24:02
		VGVALAVF	H-2-Dd, H-2-Ld	HLA-B*58:01, HLA-B*15:01, HLA-B*57:01, HLA-A*23:01, HLA-A*24:02, HLA-B*35:01, HLA-B*53:01, HLA-B*51:01
		GWLIVGVAL	H-2-Kd, H-2-Kd	HLA-A*23:01, HLA-A*24:02
		DFVRATATI	H-2-Kd	HLA-A*24:02, HLA-A*23:01, HLA-B*51:01, HLA-B*08:01, HLA-A*68:02
		GVEHVTFEY	Not Available	HLA-A*01:01, HLA-B*44:03, HLA-B*44:02, HLA-A*30:02, HLA-A*26:01, HLA-A*11:01, HLA-B*15:01, HLA-B*40:01
		RIFTIGTVTL	Not Available	HLA-A*32:01, HLA-A*02:03, HLA-A*02:06, HLA-A*02:01, HLA-A*24:02, HLA-A*23:01, HLA-B*15:01, HLA-A*30:01, HLA-B*07:02, HLA-A*68:02, HLA-B*57:01, HLA-B*58:01, HLA-A*30:02, HLA-A*03:01

NSP3	95.11%	TAFGLVAEW	H-2-Dd, H-2-Ld	HLA-B*58:01, HLA-B*53:01, HLA-B*57:01, HLA-A*32:01, HLA-B*35:01, HLA-A*26:01, HLA-B*51:01, HLA-B*44:03, HLA-B*44:02, HLA-A*23:01, HLA-A*24:02, HLA-A*68:02, HLA-B*15:01, HLA-A*30:02, HLA-A*68:01, HLA-A*01:01, HLA-A*33:01
		SYLFQHANL	H-2-Dd, H-2-Kd, H-2-Db, H-2-Kb	HLA-A*24:02, HLA-A*23:01, HLA-B*08:01, HLA-A*32:01, HLA-A*30:01, HLA-A*33:01, HLA-A*31:01, HLA-A*30:02
		LVSDIDITF	H-2-Ld	HLA-B*35:01, HLA-B*53:01, HLA-A*32:01, HLA-B*58:01, HLA-B*15:01, HLA-A*26:01, HLA-B*57:01, HLA-A*23:01, HLA-A*02:06, HLA-A*24:02, HLA-A*68:02, HLA-A*30:02, HLA-A*02:01, HLA-A*01:01, HLA-B*51:01, HLA-B*44:02, HLA-B*44:03, HLA-A*68:01, HLA-B*40:01, HLA-B*07:02, HLA-A*02:03
		TPRDLGACI	H-2-Ld	HLA-B*07:02, HLA-B*51:01, HLA-B*53:01, HLA-B*35:01, HLA-B*08:01
		EAFEYYHTT	H-2-Kb	HLA-A*68:02, HLA-B*51:01, HLA-A*26:01, HLA-B*35:01, HLA-B*08:01,

				HLA-B*53:01, HLA-A*33:01, HLA-A*02:06
NSP12	91.62%	VYANLGERV	H-2-Dd, H-2-Kd, H-2-Kb	HLA-A*23:01, HLA-A*24:02
		SSVELKHFFF	H-2-Dd, H-2-Ld, H-2-Db	HLA-B*57:01, HLA-B*58:01, HLA-A*32:01, HLA-A*01:01, HLA-A*23:01, HLA-A*26:01, HLA-A*24:02, HLA-B*53:01, HLA-B*15:01, HLA-A*30:02
		NPHLMGWDY	H-2-Ld	HLA-A*01:01, HLA-A*26:01, HLA-B*35:01, HLA-B*53:01, HLA-B*07:02, HLA-B*51:01
		VPFVSTGY	H-2-Ld	HLA-B*35:01, HLA-B*53:01, HLA-A*26:01, HLA-B*51:01, HLA-B*07:02, HLA-A*30:02, HLA-A*01:01, HLA-B*15:01, HLA-B*44:03, HLA-A*68:01, HLA-B*44:02, HLA-B*58:01, HLA-A*32:01
		AMRNAGIVG V	Not Available	HLA-A*02:03, HLA-A*02:01, HLA-A*02:06, HLA-A*30:01

Table 3. Top common immunogenic nsp3, nsp12 and ORF3a CD8+ T cell epitopes to all four variants of concern in SARS-CoV-2 and their population coverage. These epitopes are antigenic, non-allergenic,

non-toxic and stable. Epitopes that are presented on murine MHC were prioritized for selection to population coverage analysis. No ORF9b-derived top epitopes were found to have either a weak or strong binding affinity to a mouse MHC.

Protein	Population Coverage	Epitope	HLA-II restriction
ORF3a	99.76%	GVHFVCNLLLLFVTV	HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*02:01/DPB1*05:01
		FVTVYSHLLLVAAGL	HLA-DRB1*01:01, HLA-DPA1*03:01/DPB1*04:02, HLA-DQA1*01:02/DQB1*06:02, HLA-DRB1*07:01, HLA-DPA1*01:03/DPB1*04:01, HLA-DRB1*15:01
		VHFVCNLLLLFVTVY	HLA-DPA1*03:01/DPB1*04:02, HLA-DPA1*01:03/DPB1*02:01, HLA-DPA1*02:01/DPB1*01:01, HLA-DPA1*02:01/DPB1*05:01
		HLLLVAAGLEAPFLY	HLA-DQA1*05:01/DQB1*02:01
		SHLLLVAAGLEAPFL	HLA-DQA1*05:01/DQB1*02:01
		PFGWLIVGVALLAVF	HLA-DPA1*03:01/DPB1*04:02, HLA-DRB1*01:01, HLA-DQA1*05:01/DQB1*02:01, HLA-DRB1*08:02, HLA-DQA1*01:02/DQB1*06:02, HLA-DQA1*05:01/DQB1*03:01
		LLLVAAGLEAPFLYL	HLA-DQA1*05:01/DQB1*02:01, HLA-DQA1*05:01/DQB1*03:01
NSP3	99.82%	LTAFGGLVAEWFLAYI	HLA-DPA1*01:03, HLA-DPB1*02:01, HLA-DQA1*05:01, HLA-DQB1*02:01, HLA-DPA1*01:03, HLA-DPB1*04:01, HLA-DQA1*01:01, HLA-DQB1*05:01, HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DPA1*02:01, HLA-DPB1*01:01, HLA-DQA1*03:01, HLA-DQB1*03:02
		PTVVVNAANVYLKHG	HLA-DRB1*13:02, HLA-DRB1*08:02, HLA-DQA1*01:02, HLA-DQB1*06:02

		DLTAFGLVAEWFLAY	HLA-DPA1*01:03, HLA-DPB1*02:01, HLA-DQA1*05:01, HLA-DQB1*02:01, HLA-DPA1*01:03, HLA-DPB1*04:01, HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DQA1*03:01, HLA-DQB1*03:02, HLA-DQA1*04:01, HLA-DQB1*04:02, HLA-DQA1*01:01, HLA-DQB1*05:01, HLA-DPA1*02:01, HLA-DPB1*01:01
NSP12	96.62%	NLKYAISAKNRARTV	HLA-DPA1*02:01, HLA-DPB1*14:01, HLA-DRB1*11:01, HLA-DRB1*09:01
		LKYAISAKNRARTVA	HLA-DPA1*02:01, HLA-DPB1*14:01, HLA-DRB1*11:01
		PDILRVYANLGERVR	HLA-DRB1*15:01, HLA-DRB1*13:02, HLA-DRB1*04:05
		RVYANLGERVRQALL	HLA-DRB3*02:02
		GLVASIKNFKSVLYY	HLA-DRB1*15:01, HLA-DRB1*12:01, HLA-DRB1*04:01, HLA-DRB1*08:02
		VASIKNFKSVLYYQN	HLA-DRB1*15:01, HLA-DRB1*12:01, HLA-DPA1*02:01, HLA-DPB1*05:01, HLA-DRB1*08:02, HLA-DPA1*01:03, HLA-DPB1*04:01, HLA-DRB1*04:01, HLA-DRB1*04:05
ORF9b	94.74%	KAFQLTPIAVQMTKL	HLA-DPA1*02:01, HLA-DPB1*14:01, HLA-DRB4*01:01, HLA-DRB1*04:01, HLA-DRB1*01:01
		LEDKAFQLTPIAVQM	HLA-DPA1*02:01, HLA-DPB1*14:01, HLA-DPA1*01:03, HLA-DPB1*04:01
		SLEDKAFQLTPIAVQ	HLA-DPA1*02:01, HLA-DPB1*14:01
		NSLEDKAFQLTPIAV	HLA-DQA1*01:01, HLA-DQB1*05:01

Table 4. Top common antigenic CD4+ T cell epitopes across VOC and their population coverage. The selected epitopes are IFN- γ inducing, non-allergenic, non-toxic and stable.

Protein	Population Coverage	CD4+ T cell epitope	CD8+ T cell epitope	HLA-II alleles	HLA-I alleles
ORF3a	100.0%	GVHFVCNLLLLFVTV	NLLLLFVTV	HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DPA1*02:01, HLA-DPB1*01:01, HLA-DPA1*01:03, HLA-DPB1*02:01, HLA-DPA1*02:01, HLA-DPB1*05:01	HLA-A*02:01, HLA-A*02:06, HLA-A*02:03, HLA-B*08:01, HLA-A*68:02, HLA-A*32:01, HLA-B*51:01
		FVTVYSHLLLVAAGL	HLLLVAAGL	HLA-DRB1*01:01, HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DQA1*01:02, HLA-DQB1*06:02, HLA-DRB1*07:01, HLA-DPA1*01:03, HLA-DPB1*04:01, HLA-DRB1*15:01, HLA-DRB5*01:01	HLA-A*02:01, HLA-A*02:03, HLA-A*02:06, HLA-B*08:01, HLA-A*32:01
		VHFVCNLLLLFVTVY	NLLLLFVTV	HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DPA1*01:03, HLA-DPB1*02:01, HLA-DPA1*02:01, HLA-DPB1*01:01, HLA-DPA1*02:01, HLA-DPB1*05:01	HLA-A*02:01, HLA-A*02:06, HLA-A*02:03, HLA-B*08:01, HLA-A*68:02, HLA-A*32:01, HLA-B*51:01
		HLLLVAAGLEAPFLY	LLLVAAGLE A, LLVAAGLEA , HLLLVAAGL	HLA-DQA1*05:01, HLA-DQB1*02:01	HLA-A*02:03, HLA-A*02:01, HLA-A*02:06, HLA-B*08:01, HLA-A*32:01
		SHLLLVAAGLEAPFL	LLLVAAGLE A, LLVAAGLEA , HLLLVAAGL	HLA-DQA1*05:01, HLA-DQB1*02:01	HLA-A*02:03, HLA-A*02:01, HLA-A*02:06, HLA-B*08:01, HLA-A*32:01
		PFGWLIVGVALLAVF	GWLIVGVA LL, GWLIVGVA L, LIVGVALLA V, IVGVALLAV	HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DRB1*01:01, HLA-DQA1*05:01, HLA-DQB1*02:01, HLA-DRB1*08:02, HLA-DQA1*01:02, HLA-DQB1*06:02, HLA-DQA1*05:01, HLA-DQB1*03:01	HLA-A*02:01, HLA-A*02:03, HLA-A*23:01, HLA-A*24:02, HLA-A*02:06, HLA-A*68:02, HLA-A*32:01, HLA-B*51:01

	LLLVAAGLEAPFLYL	LLLVAAGLE A, LLVAAGLEA , HLLLVAAGL	HLA- DQA1*05:01/DQB1 *02:01, HLA- DQA1*05:01/DQB1 *03:01	HLA-A*02:03, HLA-A*02:01, HLA-A*02:06, HLA-B*08:01, HLA-A*32:01
NSP3	LTAFGLVAEWFLAYI	LTAFGLVAE W, TAFGLVAE W	HLA-DPA1*01:03, HLA-DPB1*02:01, HLA-DQA1*05:01, HLA-DQB1*02:01, HLA-DPA1*01:03, HLA-DPB1*04:01, HLA-DQA1*01:01, HLA-DQB1*05:01, HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DPA1*02:01, HLA-DPB1*01:01, HLA-DQA1*03:01, HLA-DQB1*03:02	HLA-B*57:01, HLA-B*58:01, HLA-B*53:01, HLA-A*32:01, HLA-A*26:01, HLA-B*35:01, HLA-B*51:01, HLA-B*44:03, HLA-B*44:02, HLA-A*23:01, HLA-A*24:02, HLA-A*68:02, HLA-B*15:01, HLA-A*30:02, HLA-A*68:01, HLA-A*01:01, HLA-A*33:01
	PTVVVNAANVYLKH G	VVVNAANV Y	HLA-DRB1*13:02, HLA-DRB3*02:02, HLA-DRB1*08:02, HLA-DQA1*01:02, HLA-DQB1*06:02	HLA-A*30:02, HLA-A*26:01, HLA-B*35:01, HLA-A*01:01, HLA-B*53:01, HLA-A*32:01, HLA-A*11:01, HLA-B*58:01, HLA-B*57:01, HLA-A*03:01, HLA-A*68:01
	DLTAFGLVAEWFLA Y	LTAFGLVAE W, TAFGLVAE W	HLA-DPA1*01:03, HLA-DPB1*02:01, HLA-DQA1*05:01, HLA-DQB1*02:01, HLA-DPA1*01:03, HLA-DPB1*04:01, HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DQA1*03:01, HLA-DQB1*03:02, HLA-DQA1*04:01, HLA-DQB1*04:02, HLA-DQA1*01:01, HLA-DQB1*05:01, HLA-DPA1*02:01, HLA-DPB1*01:01	HLA-B*57:01, HLA-B*58:01, HLA-B*53:01, HLA-A*32:01, HLA-A*26:01, HLA-B*35:01, HLA-B*51:01, HLA-B*44:03, HLA-B*44:02, HLA-A*23:01, HLA-A*24:02, HLA-A*68:02, HLA-B*15:01, HLA-A*30:02, HLA-A*68:01, HLA-A*01:01, HLA-A*33:01

		TAFGLVAEWFLAYIL	TAFGLVAE W	HLA-DPA1*01:03, HLA-DPB1*02:01, HLA-DPA1*01:03, HLA-DPB1*04:01, HLA-DQA1*01:01, HLA-DQB1*05:01, HLA-DQA1*05:01, HLA-DQB1*02:01, HLA-DPA1*02:01, HLA-DPB1*01:01, HLA-DPA1*03:01, HLA-DPB1*04:02, HLA-DPA1*02:01, HLA-DPB1*05:01	HLA-B*58:01, HLA-B*53:01, HLA-B*57:01, HLA-A*32:01, HLA-B*35:01, HLA-A*26:01, HLA-B*51:01, HLA-B*44:03, HLA-B*44:02, HLA-A*23:01, HLA-A*24:02, HLA-A*68:02, HLA-B*15:01, HLA-A*30:02, HLA-A*68:01, HLA-A*01:01, HLA-A*33:01
NSP12		NLKYAISAKNRARTV	SAKNRART V	HLA-DPA1*02:01, HLA-DPB1*14:01, HLA-DRB3*02:02, HLA-DRB5*01:01, HLA-DRB1*11:01, HLA-DRB1*09:01	HLA-B*08:01, HLA-B*51:01, HLA-A*30:01, HLA-B*07:02, HLA-A*02:03, HLA-A*30:02, HLA-A*26:01
		LKYAISAKNRARTVA	SAKNRART V	HLA-DPA1*02:01, HLA-DPB1*14:01, HLA-DRB5*01:01, HLA-DRB3*02:02, HLA-DRB1*11:01	HLA-B*08:01, HLA-B*51:01, HLA-A*30:01, HLA-B*07:02, HLA-A*02:03, HLA-A*30:02, HLA-A*26:01
		PDILRVYANLGERV R	VYANLGER V	HLA-DRB3*02:02, HLA-DRB1*15:01, HLA-DRB1*13:02, HLA-DRB1*04:05	HLA-A*24:02, HLA-A*23:01
		RVYANLGERVRQAL L	VYANLGER V	HLA-DRB3*02:02	HLA-A*24:02, HLA-A*23:01
		LEDKAFQLTPIAVQM	DKAFQLTPI	HLA-DPA1*02:01, HLA-DPB1*14:01, HLA-DPA1*01:03, HLA-DPB1*04:01	HLA-B*51:01
ORF9b		SLEDKAFQLTPIAVQ	DKAFQLTPI	HLA-DPA1*02:01, HLA-DPB1*14:01	HLA-B*51:01
		NSLEDKAFQLTPIAV	DKAFQLTPI	HLA-DQA1*01:01, HLA-DQB1*05:01	HLA-B*51:01

Table 5. Top overlapping immunogenic, antigenic, non-allergenic, non-toxic, stable, IFN- γ inducing CD8 epitopes and antigenic, non-allergenic, non-toxic, stable, IFN- γ inducing CD4 epitopes, common across

VOC. Population coverage combines both MHC class I and II presented epitopes across nsp3, nsp12, ORF3a and ORF9b.