$CD8^+$ T Cell Reactivity to SARS-CoV-2

by

Jacqueline Carmona

A Thesis Presented in Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

Approved October 2023 by the

Graduate Supervisory Committee:

Karen S. Anderson, Chair
Douglas Lake
Marco Mangone
Carlo Maley
Joshua LaBaer

ARIZONA STATE UNIVERSITY

December 2023

ABSTRACT

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) pandemic, declared in March 2020 resulted in an unprecedented scientific effort that led to the deployment in less than a year of several vaccines to prevent severe disease, hospitalizations, and death from coronavirus disease 2019 (COVID-19). Most vaccine models focus on the production of neutralizing antibodies against the spike (S) to prevent infection. As the virus evolves, new variants emerge that evade neutralizing antibodies produced by natural infection and vaccination, while memory T cell responses are long-lasting and resilient to most of the changes found in variants of concern (VOC).

Several lines of evidence support the study of T cell-mediated immunity in SARS-CoV-2 infections. First, T cell reactivity against SARS-CoV-2 is found in both (cluster of differentiation) $CD4^+$ and $CD8^+$ T cell compartments in asymptomatic, mild, and severe recovered COVID-19 patients. Second, an early and stronger $CD8^+$ T cell response correlates with less severe COVID-19 disease [1-4]. Third, both $CD4^+$ and $CD8^+$ T cells that are reactive to SARS-CoV-2 viral antigens are found in healthy unexposed individuals suggesting that cross-reactive and conserved epitopes may be protective against infection.

The current study is focused on the T cell-mediated response, with special attention to conserved, non-spike-cross-reactive epitopes that may be protective against SARS-CoV-2. The first chapter reviews the importance of epitope prediction in understanding the T cell-mediated responses to a pathogen. The second chapter centers on the validation of SARS-CoV-2 $CD8^+$ T cell predicted peptides to find conserved, immunodominant, and immunoprevalent epitopes that can be incorporated into the next generation of vaccines against severe COVID-19 disease. The third chapter explores pre-existing immunity to SARS-CoV-2 in a pre-pandemic cohort and finds two highly immunogenic epitopes that are conserved among human common cold coronaviruses (HCoVs). To end, the fourth

chapter explores the concept of T cell receptor (TCR) cross-reactivity by isolating SARS-CoV-2-reactive TCRs to elucidate the mechanisms of cross-reactivity to SARS-CoV-2 and other human coronaviruses (HCoVs).

DEDICATION

To my unofficial PI Juan, to my courageous dreamer Alo,

to my witty explorer Ana and my tenacious and bright Lucas.

You amaze and inspire me every day. I dedicate this effort to you.

"... es en ti la ilusión de cada día"

P.N

ACKNOWLEDGMENTS

I would like to acknowledge my mentor, Karen for her patience, her support, and most importantly for allowing me to join her lab. I have been privileged to find a place where I could make mistakes and grow. Thank you for all you have taught me and for pushing me to be better. Dr. Maley thank you for your kindness and advice, your work and insight always direct my thinking in new ways. Dr Mangone, thank you for pushing me forward I don't know how you always came to ask me for updates at the right moment that was needed, you taught me the value of remembering we are all the same and that I have something to contribute too. Dr. Lake, you are always very busy, but you never refuse a meeting or avoid asking questions that can help my research, thank you for your genuine interest. Dr La Baer, thank you for your questions to keep me focused and thinking about the whys and the directions of my research, CDP meetings and discussions were always fun when you were around.

Thank you to the beautiful ladies I met during my stint as a clinical RA. Your words of encouragement have never left me. I hope that my future work can honor my friend Maria Auxiliadora "Auchi" Puig, I miss your laugh, your caramel flan, and your swearing. You encouraged me to leave my country to pursue my dream of science. Every day, before I enter Biodesign, I pause and remember why I am here. The burden of cancer and disease in my family has been heavy and life-changing, it gives me purpose and humbles me.

I want to celebrate all the people in this building who have provided help and guidance. Peaches, you were the first to take me under your wing by showing me the tricks to heat agarose, bringing me food when I was too distracted to do anything else. Emma, you gave me the confidence to make mistakes, and to be unapologetic for trying, and you helped me relax and enjoy the work. Sandy, thank you for your generosity and spark, I am glad I get to see you still. Padhma and Marika, thank you for all you have done to help me through

the many projects and the perspective and insight to move forward. Thank you, Eric, for the science conversations and encouragement during the last two years, I can't wait to see what you will do next! A special thank you to Femina for her mentorship, incredible work ethic, and intelligence, I miss you! A big thank you to my friends Stacy and Anelia, you are both incredibly smart and talented, and I am lucky to know you both. Thank you Siril for your kindness and positive outlook, and to Mark, Skylar, and Dalton for the rants and laughter. Heather, Kavita, and Teresa, I hope I can pay you back your kindness during comps and after. A big thank you to my undergrads, Maggie, Deon, and Brendan, what an incredible privilege to mentor such intelligent and motivated young scientists. Finally, a big thank you to all the members of our lab, past and present as well as the members of the genomics core, Joy, Dee, Juan, and Christina; members of the La Baer lab (Dustin G, Lusheng, Lillian, Lily) as well as Dr. Murugan for your help and guiding questions. The work we do requires collaboration and kindness to each other, I appreciate all of you.

To end, thank you to my family. My dad Zider taught me the value of books and music. Mom, you always show me the beauty in all things through your paintings and creativity. My beautiful, smart, and talented sister Lina, and her three boys Michal, David, and Sebastian. My kind and generous brother Felipe, Natalia, and the sparklers Paulina and Martina. My partner in coffee and photography, my big brother Diego. To my brother Carlos, Marcela and Juanes I think of you in this new journey. My uncle Elmer, I hope we get to play and sing together again. My in-laws Guimer and Yolanda, thank you for the advice and sense of humor. Juan, without you this effort would not have been possible, too much to say for the short space I have. You keep me going, you make me laugh, you challenge me, you make me better. Alejandro, Ana Maria, and Lucas, you are my miracle. Everything I do is for you.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

CD8$^+$ T Cell Epitope Discovery

And its Use in The Characterization of The Adaptive T-Cell Response To Infectious Disease

T cells interrogate the surface of nucleated or professional antigen-presenting cells for antigens derived from pathogens or abnormal proteins expressed by cancer cells. Specifically, T cells use their T cell receptor (TCR) to probe antigens in a complex with the major histocompatibility complex (MHC) molecule on the surface of the target cell [5] (Fig. 1). The antigens are the product of protein degradation by two main pathways, depending on the source of the protein. If the source is extracellular, then the antigenic peptides are generated by proteolysis of the pathogen in endosomes, phagosomes, or lysosomes. The peptides that range between 13-17 residues are then loaded onto MHC-II molecules and presented on the cell surface to be probed by CD4$^+$ helper T cells. On the other hand, misfolded or viral proteins that originate intracellularly, are degraded in the cytosol by the constitutive proteasome or the immunoproteasome and further cleaved in the endoplasmic reticulum (ER) to a peptide between 8-10 amino acids long (Fig. 2). The shorter peptides are loaded onto the closed peptide binding groove of the MHC-I molecule and presented on the cell surface for recognition by CD8$^+$ T cells. In general, peptides that are recognized in complex with the MHC molecule via interaction with the TCR, and induce an immune response are known as **epitopes** [6].

One of the key challenges in the process of defining immunogenic T cell epitopes is the high interpersonal variability of peptide recognition owing to many reasons. The first is the highly polymorphic nature of the MHC molecule that has more than 13,000 different class I alleles (and > 5760 Class II) and determines the personal repertoire of peptides that

are loaded and presented on the cell surface for recognition [7]. Most of the polymorphisms affect the peptide binding groove and thus determine which peptides are accommodated and presented. The second reason is the immunological history of an individual that will shape activation of memory or de novo T cell responses, influence recognition due to cross-reactivity generated during prior infections with similar pathogens, or dampen activation due to similarity of the peptide/s to self-proteins [8, 9]. Other factors that influence immunogenicity include the abundance of the peptide:MHC complex that in turn is affected by the stability of the complex, the binding affinity of the peptide to the MHC, the immune-prevalence of the source protein as well as the efficiency of proteasome degradation and ER-mediated peptide cleaving [10-15]. In all, epitope recognition is defined by individual genetic and environmental factors that complicate T-cell epitope mapping efforts.

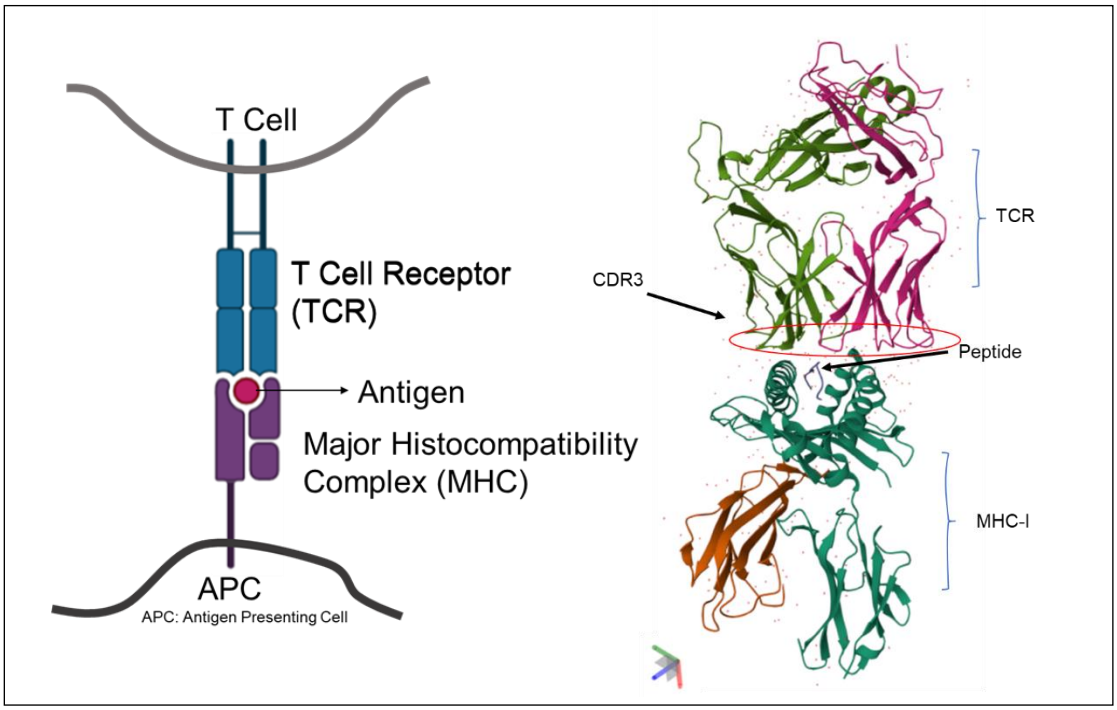


Figure 1. pMHC:TCR complex. The peptide in grey in complex with the MHC in teal and the TCR in lime green and light magenta. 3D Model made with Mol, PDB 2VLJ [16].

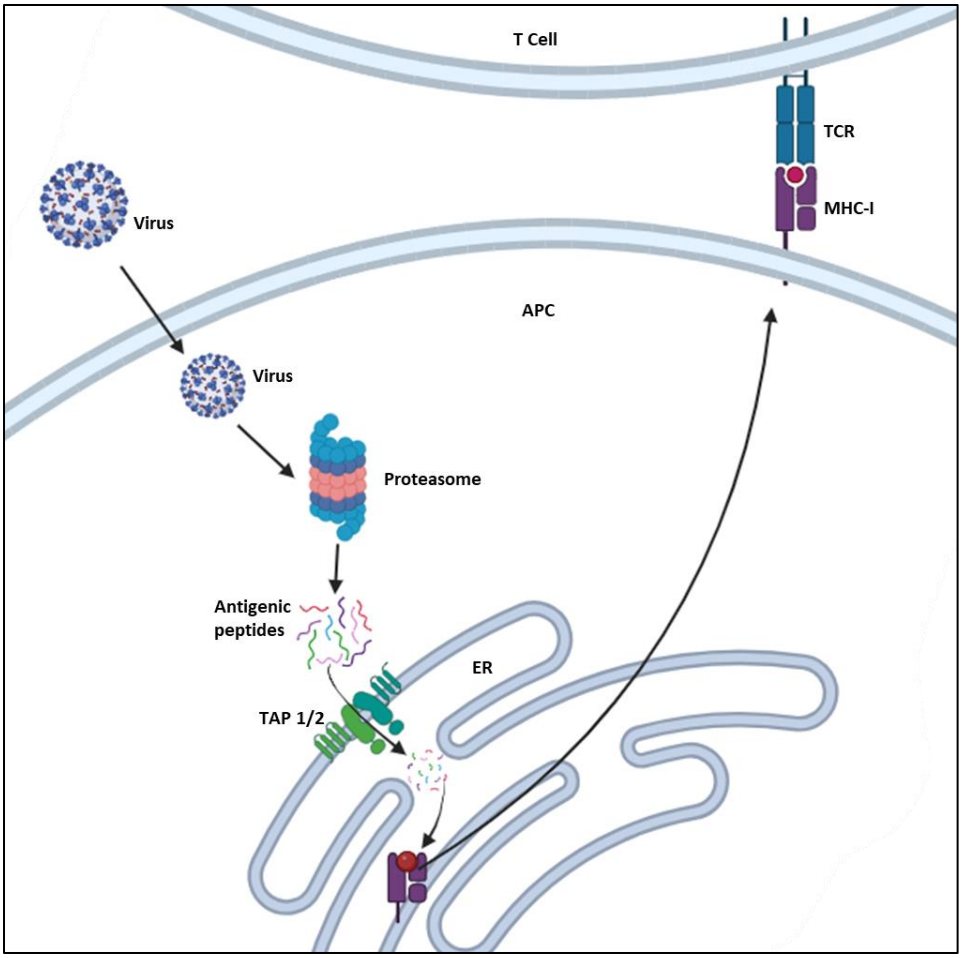


Figure 2. Antigen presentation Pathway. Adapted from ref 1. Antigens derived from viruses or misfolded proteins are degraded by the proteasome, or in the case of the immune response the immunoproteasome. The resulting peptides are translocated into the endoplasmic reticulum (ER) lumen through the transporter associated with antigen processing (TAP 1/2). In the ER, peptides may be processed further by the ER aminopeptidase-associated antigen processing (ERAAP) and then loaded onto the MHC class I molecule. The pMHC complex is then shuttled to the cell surface for interrogation by $CD8^+$ T cells. LMP2: Low molecular weight polypeptide complex. TCR: T cell receptor. APC: Antigen Presenting Cell. Figure made with BioRender.

1.1 HLA Class I Supertypes and epitope recognition

The structural definition of more than 600 pMHC complexes by X-ray crystallography has provided insight into the biochemical parameters involved in allele-specific binding patterns to a specific peptide [17]. Although, most of the HLA class I allele structures have not been resolved, there are common patterns that rule peptide and MHC molecule

interaction. For instance, HLA class I alleles have six discrete binding pockets within their peptide binding groove (Fig. 3) named with the letters A through F. Each binding pocket has an allele-specific biochemical structure that determines which and how each residue from the peptide is accommodated. Pockets B and F interact with anchor positions 2 (P2) and the C-terminal (PΩ) amino acid of the peptide, respectively [18-21]. Anchor residues take their name from the central role they have in binding to the B and F pockets and anchoring the peptide within the peptide binding groove [20]. Guided by the nature of the P2 and P9 residues, Sidney et al. classified HLA alleles into supertypes or groups of alleles that share chemical properties or motifs, at the B and F pockets (Table 1). Using this grouping, they defined 9 different HLA supertypes that fully and partially match about 57% and 23.8% respectively, of most HLA class I alleles [19]. Supertype classification has facilitated the mapping of putative MHC class I ligands while identification of overlapping peptide repertoires helps to project population coverage.

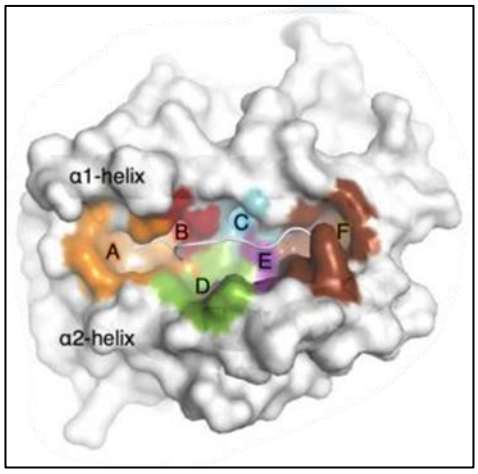


Figure 3. Representation of the peptide binding groove of the HLA class I molecule. Each of the binding pockets or clefts is represented with a color. Pocket A: Orange, B: Red, C: Light blue, D: green, E: purple, F: brown. The peptide is represented in white. Modified from Ref 21.

Table 1. B and F pocket amino acid affinity according to HLA class I supertype. Modified from reference [20].

Supertype	B Pocket Residues	F pocket Residues
A01	Small/Aliphatic	Aromatic/Large/Hydrophobic
A01/A03	Small/Aliphatic	Aromatic/Basic
A01/A24	Small/Aliphatic/Aromatic	Aromatic/Large/Hydrophobic
A02	Small/Aliphatic	Aliphatic/Small/Hydrophobic
A03	Small/Aliphatic	Basic
A24	Aromatic/Aliphatic	Aromatic/Large/Hydrophobic
B07	Proline	Aromatic/Large/Hydrophobic
B08	Undefined	Aromatic/Large/Hydrophobic
B27	Basic	Aromatic/Large/Hydrophobic/Basic
B44	Acidic	Aromatic/Large/Hydrophobic
B58	Small	Aromatic/Large/Hydrophobic
B62	Aliphatic	Aromatic/Large/Hydrophobic

A few models have tried to predict which residues or combinations of peptides facing the TCR, are likely to induce T cell reactivity. Since peptide residues P4-P6 are involved in the contact with TCR residues, Calis et al, devised a model that looks at the hydrophobicity of non-anchor residues and found that immunogenic epitopes tend to be more hydrophobic than non-immunogenic peptides. This model can explain 65% of observed immunogenicity and it is available as a free open-access platform at http://tools.iedb.org/immunogenicity/ [22]. A similar study confirmed that TCR contact residues in both human and murine models have a higher hydrophobicity index and lower polarity among 9-mer epitopes restricted to HLA-A*02:01 and MHC H-2D^{b} and H-2K^{b} [23]. More recent platforms like pMTnet, incorporate complementary determining region 3 (CDR3β) features from the TCR with the peptide and HLA class I sequences, into a deep-learning network. pMTnet is trained on TCR:pMHC binding/not-binding pairs to identify immunogenic interaction [24]. However, the amino acid sequence of the peptide and binding affinity to the MHC complex alone are not enough to define immunogenicity. Integrating the biochemical character of each amino acid, especially residues P4-P6 and

sometimes P7, and binding elements of the TCR (CDR3) can improve current predictive models of reactivity which can be useful for vaccine and immunotherapy design.

1.2 Immunodominance and immunoprevalence

An important variable in epitope identification is the abundance of target peptide and how it shapes the TCR repertoire during an immune response. Epitope recognition that elicits a higher magnitude of response than other epitope-specific responses is known as **immunodominant [25]**. Several factors can contribute to immunodominance: the abundance of the source protein, the expression level of the HLA to which the epitope is restricted, and the efficiency of processing and presentation of the antigenic protein [14]. Differences in cleavage patterns by the immunoproteasome influence the $CD8^+$ T cell repertoire, for example favoring cleaving at the PΩ position of hydrophobic residues. However, incorporating these factors in a model to predict cleaving and presentation has been modestly successful, missing about half of the possible peptides (reviewed in [26]). Part of the reason is that not all the rules of the game are clear; mice deficient in all three subunits of the immunoproteasome are severely but not completely impaired at antigen processing and presentation. When wild-type (WT) splenocytes are transferred into triple knock out (TKO) mice, the cells are rejected, indicating that the type of peptides processed and presented are different in TKO mice. Only 50% of peptides presented were similar between WT and the TKO mice. In addition, T cells from these mice are unable to replicate and survive when transferred into wild-type-virus-infected mice, suggesting a possible role for the immunoproteasome in T cell survival and proliferation during infection [27].

In addition to immunodominance, which is defined by the magnitude of response, immunoprevalence is also central for learning about T cell reactivity and the breadth of the T cell response. **Immunoprevalence** is the frequency at which an epitope is recognized by T cells, regardless of the magnitude. Many times, immunoprevalent

epitopes are also immunodominant mainly because abundance influences both. However, immunodominance is more dependent on individual history, TCR repertoire, and HLA restriction while prevalence is defined by immunogenic areas of an antigen instead of a single peptide. These areas can be identified by different individuals with different HLA restrictions, as in the case of pox virus open reading frames (ORF) [28]. For smaller viruses like HIV and Hepatitis C, immunoprevalence is harder to distinguish from immunodominance since the repertoire of processed proteins is smaller and more likely to be important throughout their life cycle. In contrast, larger viruses like the herpes virus EBV (Epstein Bar Virus) and VV (vaccinia virus) have a $CD8^+$ T cell immunoprevalent hierarchy were peptides from larger proteins may be slightly immunogenic or not at all. About 50% of VV proteins encode $CD8^+$ T cell epitopes and 10 epitopes originate from the same antigenic protein and are also restricted to 7 different HLAs [28]. Oseroff et al. propose that immunoprevalence may be more dependent on the abundance of the protein, cellular localization, and patterns of expression and degradation.

Identification of immunodominant epitopes ensures identification of immunogenic targets according to abundance and magnitude of response while understanding immunoprevalence can help to expand coverage of vaccines and treatments to a wider set of HLA restrictions. Incorporating immunodominant and immunoprevalent information into predictive algorithms can improve epitope discovery from larger proteins that may not be cost/effort-effective to screen in their entirety.

1.3 MHC ligand and Immunogenicity predictions

The identification of epitopes, is one of the many steps required to understand an immune response against a pathogen or a nascent tumor and it is dependent on the host, the source of the antigen (virus, tumor), and whether the source of the reaction is via vaccination, natural infection or in vitro/ex vivo stimulation. In general, we can

differentiate peptides from epitopes by the combination of three main types of tests. The first is by mass spectrometry of eluted MHC-bound ligands (ELs), this method indirectly provides information on cleavage and presentation repertoires but not immunogenicity [29]. The second assay measures MHC binding in vitro to quantify binding affinity (BA) and HLA restriction, however, it does not provide data on natural antigen processing and presentation [30]. The third type of assay measures T cell epitope recognition and provides information on the magnitude and polarity of the T cell response but it may be biased to the HLAs being tested and miss-represent the actual immunopeptidome [31]. Due to the complexity and labor-intensive nature of theses assays, computational tools have been developed to aid in the effort to determine the immunopeptidome of a specific pathogen or tumor.

The last three decades have demarcated the birth and surge of computational models aimed at predicting the peptide repertoire of MHC class I and II alleles. Most algorithms provide information on predicted BA as IC_{50} values between the peptide and the MHC molecule. IC50 is defined as the concentration in nM at which 50% of binding of a labeled reference peptide is inhibited and it approximates the equilibrium dissociation constant (K_D) between the peptide and the respective MHC molecule. The lower the IC_{50} value, the higher the binding affinity [32, 33]. Recently, several algorithms have been developed that link machine learning (e.g. SMM) [34], advanced neural networks (e.g. NetMHC3.2) [35], and optimally connected Hidden Markov Models (ocHMM) with experimentally derived datasets on HLA binding affinity, processing and cleaving [6, 36-38]. Models that integrate ligand elution and binding affinity data can predict >90% of naturally presented peptides with 98% specificity [39]. However, the number of truly immunogenic peptides is much lower, between 10-15% when using most predictive algorithms, and even lower for neo-antigen epitope discovery (<5%) [40].

Epitope discovery using individual predictive algorithms provides results that are biased to the method and dataset used to train the algorithm, e.g. limited to certain HLA restrictions, viral sequences, representative of a few ethnicities, etc. Pooling multiple epitope prediction algorithms and benchmarking with experimentally validated sets of eluted peptides (ELs) and binding affinities (BA) has been shown to increase the positive predictive value (PPV) of MHCs ligand prediction as well as the percentage of truly immunogenic epitopes [35, 40, 41]. For instance, Lundegaard et al, modified the NetMHC platform by using a consensus method that combines eluted and binding affinity information (NetMHC*pan-4*), to improve the percentage of discovered immunogenic peptides to 54%. On the other hand, the addition of peptide processing by the immunoproteasome and presentation on the cell surface have added minor improvements to MHC ligand predictions. The reason may be that cleavage by the immunoproteasome and TAP transport events are less selective and more conserved than MHC peptide binding. For example, the immunoproteasome tends to favor cleaving at hydrophobic residues at the C-terminus of the peptide, independent of the HLA to which it will be loaded [26]. Adding processing and presentation probabilities, although minor, does improve prediction algorithms and has extended the predicted repertoire length to include 8, 9, 10, and 11 mer peptides [42]. Overall, predictive algorithms that train on binding affinity and eluted peptides, reduce the number of putative ligands to test and streamline the process of finding the phenotype and functional profile of epitope-specific responses under different contexts, like vaccination, acute infection, convalescence, and cancer.

EnsembleMHC and the binding capacity of common HLA class I alleles to SARS-CoV-2 $CD8^+$ T cell predicted peptides

EnsembleMHC is a consensus platform that integrates predictions of binding affinity, processing, and presentation with mass spectrometry of eluted peptides and binding

affinity data [29, 43]. Specifically, it integrates 7 different algorithms: MHCFlurry-affinity-1.6.0 [44], MHCFlurry-presentation-1.6.0 [44], netMHC-4. [45], netMHCpan-4.0-EL [39], netMHCstabpan-1.0 [46], PickPocket-1.1 [47] and MixMHCpred-2.0.2 [48]. To determine allele- and algorithm-specific binding affinity thresholds, each algorithm was trained on a data set of mass-spectrometry-confirmed ELs that were naturally presented by 92 different stable monoallelic cell lines [29]. A subset of the target ELs was used, restricted to 52 of the 95 HLA class I A, B and C alleles supported by all 7 algorithms (Fig. 4).

To determine the false discovery rate (FDR), each platform was fed a set of decoy peptides (100-fold excess of the allele-specific set) randomly selected from a pool of predicted but not detected peptides from the same protein source. Then, the target and decoy ELs were ranked according to binding affinity. Allele- and algorithm-specific thresholds were defined so that 50% of the total repertoire was selected [43]. The final outputs for every algorithm and allele set, are then combined into a composite score that allows to select of the final peptide set. To do this, algorithm-specific FDRs are calculated as the fraction of target-to-decoy peptides at the specific BA threshold. Then, a composite FDR, $peptide^{FDR}$, is assigned by multiplying the 7 different empirical FDRs of each peptide. This composite score is then used to select peptides with a $peptide^{FDR}$ that is equal to or less than 5%. Selecting this level of FDR, although it allows for some false positives in the output set, prevents false positives from exceeding 5% of the total predicted repertoire [49].

Wilson et al, found that by using a composite $peptide^{FDR}$, EnsembleMHC can recall high-quality predicted peptides with higher precision than individual algorithms. Precision and recall were calculated under a combination of permissive or restrictive thresholds for each of the 7 algorithms and EnMHC. A permissive limit includes 20% of

the highest binders (binding percentile ≤2%, binding affinity ≤500nM, presentation ≤2%, FDR ≤50%) while restrictive conditions include only the top 5% (binding percentile ≤0.5%, binding affinity ≤50nM, presentation ≤0.5%, FDR ≤5%). Although EnMHC had lower recall ability both under restrictive and permissive thresholds, the positive predictive value (PPV), or precision, was 3.4-fold higher for EnMHC than for individual platforms (Fig. 4).

The setup of this algorithm allows for the identification of high binders with a high probability of being processed and presented on the cell surface with a lower FDR by using the peptide^{FDR} as a metric for discerning the top binders. In addition, it considers the different binding capacities of each HLA, instead of using a single threshold for all the algorithms in the consensus platform. EnsembleMHC successfully prioritized peptides that were experimentally validated to be immunogenic across the Hepatitis C genome polyprotein (P26664), the Dengue virus genome polyprotein (P14340), and the HIV-1 POL-GAG protein (P03369). Ensemble MHC had 1.5 to 7-fold higher call precision [True-Positives/(TP+FalsePositives)] for top-ranking viral peptides when compared to individual algorithms (e.g. netMHCpan-EL, PickPocket, etc.) [43]. Demonstrating again that pooling multiple epitope prediction algorithms improves precision and lowers FDR of predicted peptides while increasing the likelihood of finding immunogenic epitopes [39].

Peptide repertoire predictions for multiple alleles are used to represent the HLA diversity in a study population. However, different HLA class I alleles have different binding capacities or repertoires as well as different epitope binding affinities as shown in a study by Sette et al with the Dengue Virus proteome [50]. Likewise, the predicted peptide repertoire for the SARS-CoV-2 proteome using EnMHC, revealed the same trend where there is a significant difference in the binding capacity of HLA class I alleles to the SARS-

CoV-2 peptidome. After analysis with EnMHC, 3.49 million peptides were predicted and restricted to 52 of the HLA class I A, B, and C alleles supported by the consensus algorithm. After filtering out predicted peptides with a peptide^{FDR} higher than 5%, the number of total epitopes was refined from 3.49 million to 658 peptides from non-structural (nsp; n=542) and structural proteins (n=108) (Table 2) [43].

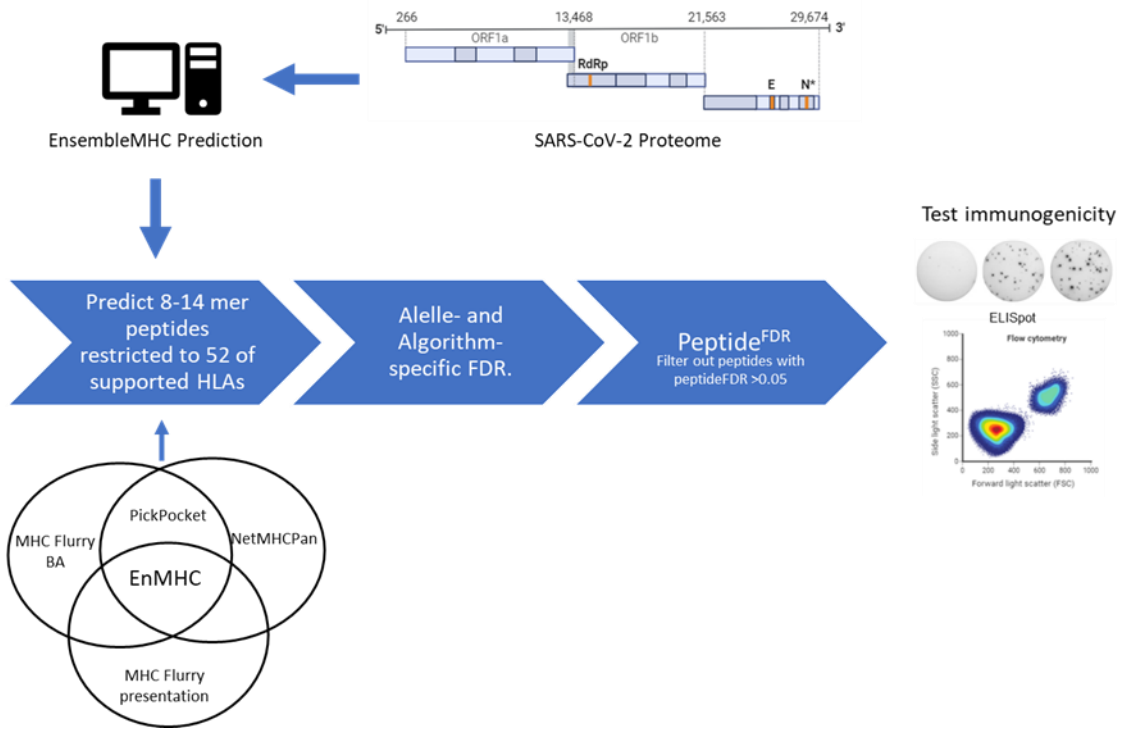


Figure 4. Schema and workflow to define high-quality predicted peptides: high binders according to IC50 and percentile, top percentile for cleavage and presentation. Peptides with a peptideFDR higher than 5% were eliminated. Adapted from Wilson et al. 2022. Illustration made with Biorender.

Peptides were distributed among 22 A, 21 B, and 9 C class I HLA alleles. HLA-A alleles had the highest binding capacity above B and C. However, one must take into account that one, C alleles are characterized by lower transcript and expression levels on the cell surface, and two, HLA-C BA and EL data is not as abundant as for the other class I alleles

[51]. The alleles with the least number of predicted epitopes were A*02:05 (2 nsp/1 structural) and B*07:02 (3 nsp/1 structural) while A*24:02, B*57:01 and B*53:01 had the highest binding capacities (see Table 3). Although the study does not look at aggregated allele-specific binding affinity, it finds that the allele-specific epitope load is significantly different when looking at the entire proteome and when comparing non-structural proteins (nsp) with structural viral proteins like Spike (S), Nucleocapsid (N), membrane (M) and Envelope (E) (Table 3) [43]. Although the inter-allele differences in binding capacities are observed both for nsp and structural proteins, the difference is more pronounced for structural proteins. The significant deviation from an even distribution of peptides among the different HLA class I alleles agrees with several studies that have observed the same phenomena when looking into $CD4^+$ and $CD8^+$ T cell predicted peptides from SARS-CoV-2 and Dengue Virus [50, 52]. The study by Wilson et al, found that the binding capacity of a population is inversely correlated to the mortality rates due to COVID-19, even after accounting for socioeconomic and high-risk factors like BMI, and cardiovascular and pulmonary disease [43].

1.4 pMHC predictions and heterologous immunity, sequence conservation, and T cell cross-reactivity

Memory responses to previous infections can alter the kinetics and patterns of the immune response to subsequent encounters with the same or different pathogen/s [53]. Heterologous immunity can occur against parasites, bacteria, and viruses and may lead to immune responses that can be protective or pathological [54]. For example, pre-existing T cell immunity to swine-origin H1N1 influenza virus due to T_{mem} cell recognition of conserved regions was associated with protection from severe disease [55, 56]. Similarly, the sequence homology of SARS-CoV-2 to common cold coronaviruses (CCCoVs) is about 65-69% and has been shown to be associated with protection against infection in COVID-

19 contacts [57, 58]. On the other hand, conserved epitopes can also have lower recognition by T cells as it is the case of proteins that share homology to self. T cells that recognize conserved epitopes from the human proteome ought to be deleted during T cell maturation by negative selection, while T cells that react to proteins from commensal organisms are suppressed as part of a protective process called peripheral tolerance [59, 60]. Both instances are examples of reduced T cell reactivity that can impact recognition to viral pathogens and thus should be considered in the design of predictive algorithms that look for MHC ligands for vaccine or therapy design.

The use of HLA binding motifs along with binding affinity and eluted ligand databases to aid machine learning simplifies and has improved epitope discovery pipelines. We have learned that there are HLA-specific variations in binding capacity and binding affinity defined by the biochemical character and spacing of individual amino acids which in turn influence affinity to the MHC molecule and determine immunogenicity. The present work seeks to validate predictions by EnsembleMHC using SARS-CoV-2 as a model (Chapter 2). In Chapter 3, the predicted peptides will be used to describe the dynamics of the $CD8^+$ T cell response against SARS-CoV-2 in vitro for 12 months. Also, reactivity will be studied after natural infection and/or vaccination. Chapter 4 will present and analyze data about pre-existing immunity in up to 50% of individuals to a virus never encountered before that produced the first worldwide pandemic since the influenza pandemic of 1921. To end, we'll discuss the biochemical mechanisms of pre-existing immunity and cross-reactivity to SARS-CoV-2 by studying a cross-reactive T cell receptor (TCR) isolated from a pre-pandemic sample reactive to the virus.

Table 2. Structural peptides predicted with EnMHC restricted to 52 common HLA-I alleles. Non-structural peptides are reported in reference 24 (Wilson et al. 2021).

Peptide Sequence	HLA Restriction	Protein	Start Position	End Position
FLAFVVFLL	A02:01|A02:11	E	20	28
LVKPSFYVY	A30:02	E	51	59
YVYSRVKNL	C12:03	E	57	65
RVKNLNSSR	A31:01	E	61	69
SLVKPSFYV	A02:11	E	70	58
EELKKLLEQW	B44:02	M	11	21
KLLEQWNLV	A02:11	M	15	23
FAYANRNRF	C12:03	M	37	45
YANRNRFLY	A29:02	M	39	47
NRFLYIIKL	B27:05|C06:02|C07:01	M	43	51
LYIIKLIFLW	A24:02	M	46	55
IAIAMACLVGLMW	B57:01	M	80	92
MACLVGLMW	B57:01	M	84	92
SYFIASFRL	A23:01	M	94	102
SYFIASFRLF	A23:01	M	94	103
YFIASFRLF	A23:01	M	95	103
ASFRLFARTRSMW	B57:01	M	98	110
RLFARTRSMW	A32:01	M	101	110
RLFARTRSMWSF	A32:01	M	101	112
SELVIGAVIL	B40:01	M	136	145
LVIGAVILR	A66:01|A68:01	M	138	146
AVILRGHLR	A31:01	M	142	150
RIAGHHLGR	A31:01	M	150	158
VATSRTLSY	B46:01	M	170	178
ATSRTLSYY	A01:01|A30:02	M	171	179
ATSRTLSYYK	A11:01	M	171	180
RVAGDSGFAAY	A30:02	M	186	196
SSDNIALLV	C05:01	M	213	221
QRNAPRITF	B27:05	N	9	17
LPNNTASW	B53:01	N	42	49
LPNNTASWF	B53:01	N	42	50
NTASWFTAL	A68:02	N	45	53
FTALTQHGK	A68:01	N	50	58
FPRGQGVPI	B07:02|B51:01|B54:01	N	63	71
KMKDLSPRW	A32:01	N	97	105
LPYGANKDGIIW	B53:01	N	118	129
RQKRTATKAY	B15:01	N	256	265
KAYNVTQAF	A32:01|B15:03|B15:17| B46:01|B57:01|C12:03	N	263	271
AQFAPSASAF	B15:01|B15:03	N	302	311
MEVTPSGTW	B44:02|B44:03	N	319	327
TPSGTWLTY	B35:01	N	322	330
LTYTGAIKL	B15:17	N	328	336
KTFPPTEPKK	A03:01|A11:01	N	358	367
KTFPPTEPK	A03:01|A11:01|A30:01	N	358	366

Table 2. Contd.

Peptide Sequence	HLA Restriction	Protein	Start Position	End Position
LPLVSSQCV	B51:01	S	8	16
YTNSFTRGV	A68:02	S	25	33
VYYPDKVFR	A31:01	S	33	41
LPFFSNVTW	B35:01|B53:01	S	53	61
LPFFSNVTWF	B53:01	S	53	62
LPFFSNVTWFHA	B54:01	S	53	64
KRFDNPVLPF	B27:05	S	72	81
LPFNDGVYF	B35:01|B53:01	S	79	87
LPFNDGVYFA	B54:01	S	79	88
GVYFASTEK	A03:01|A11:01	S	84	92
CPFGEVFNA	B54:01	S	104	112
TLDSKTQSL	C05:01	S	104	112
NATNVVIKV	A68:02	S	117	125
GVYYHKNNK	A03:01	S	142	150
SEFRVYSSA	B45:01	S	150	158
RVYSSANNCTF	A32:01	S	153	163
VYSSANNCTF	A24:02	S	154	163
SANNCTFEY	B35:01	S	157	165
FVFKNIDGY	A25:01|A26:01	S	187	195
WTAGAAAYY	A26:01|A30:02	S	253	261
YLQPRTFLL	A02:01|A02:02|A02:03| A02:05|A02:06|A02:07| A02:11|B08:01	S	264	272
ETKCTLKSF	A25:01|A26:01	S	293	301
QPTESIVRF	B35:01|B53:01	S	321	329
VRFPNITNL	B27:05|C06:02|C07:01|C07:02	S	322	330
FPNITNLCPF	B35:01|B53:01	S	324	333
SVYAWNRKR	A31:01	S	344	352
CVADYSVLY	A26:01|A29:02	S	356	364
LYNSASFSTF	A23:01|A24:02	S	368	377
NSASFSTFK	A68:01	S	370	378
KIADYNYKL	A02:02|A02:06|A32:01	S	417	425
RLFRKSNLK	A03:01	S	449	
KSNLKPFER	A31:01	S	453	461
CYFPLQSYGF	A23:01|A24:02	S	493	492
KKFLPFQQF	B15:03	S	552	560
TEVPVAIHA	B45:01	S	613	621
HADQLTPTW	B53:01	S	620	628
VYSTGSNVF	A23:01|A24:02	S	630	638
IPIGAGICASY	B35:01	S	659	667
IPIGAGICA	B54:01	S	659	667
VASQSIIAY	B35:01	S	682	690
IPTNFTISV	B51:01	S	709	717
FTISVTTEI	A68:02	S	713	721
STECSNLLLQY	A01:01	S	741	751
NTQEVFAQV	A68:02	S	772	780

Table 2. Contd.

Peptide Sequence	HLA Restriction	Protein	Start Position	End Position
RSFIEDLLF	B15:17|B57:01	S	810	818
LTDEMIAQY	A01:01	S	860	868
IPFAMQMAY	B35:01|B53:01	S	891	899
FAMQMAYRF	B53:01	S	893	901
VLNDILSRL	A02:02|A02:03	S	971	979
AEIRASANL	B40:01	S	1011	1019
AEIRASANLA	B45:01	S	1011	1020
ASANLAATK	A11:01	S	1015	1023
FPQSAPHGV	B54:01	S	1047	1055
VTYVPAQEK	A03:01|A11:01	S	1060	1069
GTHWFVTQR	A11:01|A31:01	S	1094	1102
VYDPLQPEL	A23:01|C04:01|C05:01|C14:02	S	1132	1140
DISGINASV	A68:02	S	1163	1171
KEIDRLNEV	B40:02	S	1176	1184
QELGKYEQY	B44:02|B44:03	S	1196	1204
QYIKWPWYI	A23:01|A24:02	S	1203	1211
QYIKWPWYIW	A24:02	S	1203	1212
QYIKWPWYIWLGF	A24:02	S	1203	1213
WPWYIWLGF	B53:01	S	1207	1215
FIAGLIAIV	A02:02|A68:02	S	1215	1223

Table 3. Predicted binding capacity of HLA genotypes of pre-pandemic and ASU student donors. This table provides the predicted binding capacity for the entire proteome of SARS-CoV-2 and structural proteins of each donor.

Cohort	Donor	HLA-A		HLA-B		HLA-C		# Predicted peptides	
								Entire proteome	Structural
ASU Student	ASC-7	A*01:01:71	A*02:11:01	B*57:01:01	B*40:06:01	C*15:02:01	C*06:02:01	103	14
	ASC-8	A*01:01:01	A*03:01:01	B*57:01:01	B*44:03:01	C*16:01:01	C*06:02:01	121	18
	ASC-09	A*31:01:02	A*24:07:01	B*57:01:01	B*40:06:01	C*06:02:01	C*15:02:01	64	14
	ASC-12	A*01:01:71	A*68:01:01	B*35:03:01	B*57:01:01	C*04:01:01	C*06:02:01	122	14
	ASC-13	A*02:01:126	A*02:844	B*15:01:01	B*40:01:02	C*03:04:50	C*03:288	63	6
	ASC-25	A*33:01:01	A*24:02:01	B*14:02:01	B*08:01:01	C*07:01:31	C*08:02:01	61	12
	ASC-29	A*03:01:01	A*02:01:01	B*39:06:02	B*07:02:01	C*07:02:01	C*07:744	61	10
	ASC-61	A*03:01:01	A*02:05:01	B*49:01:01	B*07:02:01	C*07:01:31	C*07:02:01	48	12
	ASC-42	A*03:355	A*30:01:01	B*13:02:01	B*07:139:02	C*07:02:01	C*06:02:01	41	4
	ASC-44	A*30:01:01	A*68:01:02	B*13:02:01	B*40:01:02	C*03:04:01	C*06:02:01	72	8
	ASC-59	A*03:01:01	A*23:01:01	B*15:01:01	B*44:03:01	C*04:01:01	C*03:04:01	104	19
	ASC-63	A*01:01:01	A*01:01:71	B*57:01:01	B*27:05:02	C*02:02:02	C*06:02:01	88	17
	ASC-67	A*03:01:01	A*02:01:01	B*51:01:01	B*27:05:02	C*15:02:01	C*02:02:02	89	15
	ASC-79	A*31:01:02	A*02:01:01	B*51:14	NULL	C*15:02:01	C*15:09:01	43	9
	ASC-95	A*02:01:01	NULL	B*52:01:01	NULL	C*12:02:02	NULL	23	5
	ASC-99	A*02:844	A*68:01:02	B*48:01:01	B*08:01:01	C*07:01:01	C*08:01:01	47	9
	ASC-107	A*01:01:01	A*32:01:01	B*18:01:01	B*52:01:01	C*05:01:01	C*12:02:02	83	12
	ASC-121	A*02:01:01	A*24:02:01	B*15:01:01	B*44:02:01	C*03:03:01	C*05:01:01	134	18
	ASC-155	A*68:01:02	A*32:01:01	B*13:02:01	B*13:02:09	C*07:04:01	C*06:02:01	64	11
Pre-pandemic	BCP2	A*3:01	A*29:01	B*44:02	B*55:01	N/A	N/A	67	7
	BCP6	A*11:01	A*24:02	B*15:02	B*40:01	N/A	N/A	127	17
	BCP8	A*01:01:01	A*24:02:01	B*55:01:01	B*08:01:01	C*07:01:31	C*03:126	95	15
	BCP9	A*01:01:01	A*24:02:01	B*37:01:01	B*35:01:01	C*04:01:88	C*06:02:01	128	23
	BCP12	A*24:02:01	A*02:01:01	B*48:01:01	B*48:01:01	C*05:01:01	C*08:01:01	95	13
	BCP13	A*01:01:01	A*02:01:01	B*57:01:01	B*44:03:01	C*16:01:01	C*06:02:01	118	14
	BCP14	A*02:01:01	A*02:01:01	B*51:01:01	B*40:01:02	C*03:04:54	C*14:02:01	93	10
	BCP15	A*02:844	A*30:01:01	B*44:02:38	B*40:02:01	C*05:01:01	C*12:03:01	73	9
	BCP16	A*02:01:01	A*29:02:01	B*40:01:02	B*44:03:01	C*16:01:01	C*16:01:01	89	10
	BCP17	A*31:01:02	A*11:01:01	B*08:01:01	B*40:01:02	C*07:01:01	C*03:344:01	92	20
	BCP18	A*33:01:01	A*24:02:01	B*35:02:01	B*14:02:01	C*04:01:01	C*08:02:01	54	9
	BCP19	A*02:01:01	A*02:01:01	B*15:01:01	B*44:02:01	C*03:04:52	C*05:01:01	78	8

CHAPTER 2

CD8$^+$ T Cell Reactivity to SARS-CoV-2 in a Young, Healthy Cohort During the Delta and Omicron Waves

2.1 Introduction

SARS-CoV-2 has claimed the lives of more than 6.6 million individuals around the world since the start of the pandemic over 3 years ago [61]. In addition to the clinical consequences of acute SARS-CoV-2 infection, up to 30% of COVID-19 convalescent individuals present with long-COVID, also known as PASC (post-acute sequelae), a chronic condition that can impair mental and physical health for more than 30 days. Long-COVID is projected to cost the US government up to 2.6 trillion dollars or about nine thousand dollars in treatments per person per year [62]. Although both therapies and vaccines against severe COVID-19 are available, the constant evolution of the virus to evade humoral immunity, and inequality in vaccine access have kept SARS-CoV-2 infection a constant threat to global health and economic stability.

SARS-CoV-2 variants evade both natural and vaccine-induced humoral immunity. Vaccine-mediated protection against severe disease and hospitalization is lower for the Omicron and Delta VOC (variants of concern) than for variants like alpha or beta that dominated earlier in the pandemic [63]. Sera from convalescent and vaccinated patients have limited neutralization against the Omicron variant and is attributed to amino acid changes at antibody neutralization sites within the spike protein [64-67]. Omicron has 30 amino acid changes in the spike protein, which explains in part the higher rate of re-infection after Omicron convalescence and highlights the need to formulate vaccines that use conserved non-spike epitopes and enhance T cell-mediated responses that supplement a decline in humoral immunity. Variant-independent CD4$^+$ and CD8$^+$ T cell reactivity has been described against conserved structural proteins such as the Nucleocapsid and the

Membrane, in addition to non-structural proteins (NSP) like nsp12 (RdRp-RNA Dependent RNA Polymerase) and nsp3 (Papain protease). Using targets that are conserved and essential during early/late viral stages can potentially supplement variant-independent humoral and innate responses to the virus with $CD8^+$ T cell-mediated responses.

An increasing body of evidence supports adding T-cell targets to a vaccine against COVID-19. Non-human primate studies show that $CD8^+$ T cells protect against SARS-CoV-2 infection and severe disease while mice models show that both $CD4^+$ and $CD8^+$ T cell populations protect against coronaviral infections [68-70]. Moreover, IFN-γ secretion upon T-cell activation correlates to mild COVID-19 disease [71]. Although T cell reactivity to SARS-CoV-2 has been extensively studied using predicted peptide mega pools or overlapping peptides from the viral peptidome, most of the peptides are restricted to common HLA class I and II alleles. Few studies have deconvoluted the immunogenicity of individual predicted peptides to map T cell reactivity longitudinally and conservation through VOC.

Here, we measured the breadth, specificity, and persistence of $CD8^+$ T cell reactivity in both vaccinated and naturally infected young adults during the Delta and Omicron waves between 2021 and early 2022. In addition, we study the impact of the newer variants on T cell responses to predicted peptides restricted over 31 different HLA-I alleles in the context of mild or asymptomatic COVID-19 disease. Our lab previously designed a consensus predictive algorithm, EnsembleMHC, that can predict HLA class I high-affinity ligands with a low false discovery rate (FDR) [43]. Using EnsembleMHC, we found 108 peptides derived from structural proteins and restricted to 52 of the most common HLA Class I alleles. We measured immunogenicity by T cell IFN-γ secretion ELISpot to 93 out of the 108 structural peptides restricted to 31 unique HLA class I alleles predicted to cover

>99.5% of the global population. We found, that EnsembleMHC predicted correctly the binding capacity of individuals and most importantly, 82% (76/93) of screened peptides were immunogenic. The breadth and persistence of the $CD8^+$ T cell response include epitopes from the spike and other structural proteins, after vaccination reactivity is enhanced to all structural proteins and can endure up to 16 months post-vaccination.

2.2 Results

2.2.1. Description of ASU cohort to assess $CD8^+$ T cell responses to SARS-CoV-2 predicted peptides

We collected blood samples from a young (median age 25 y/o) student cohort from Arizona State University (ASU) and a heavier representation of males (14/19) (Table 4). Samples were collected over 15 months, starting in February 2021 until May 2022. At the time of recruitment, 4 donors did not have detectable levels of antibodies against the spike (Access Immunoassay CLIA certified ASU Biodesign Institute) and self-reported as never having been infected with SARS-CoV-2. One of the four seronegative students had received a vaccination with J&J 8 days before the first blood draw. An additional 5 individuals self-reported as never had been infected, all of them were vaccinated and positive for antibodies against the receptor binding domain (**RBD**) of spike. The rest of the students (n=10) were sero-reactive to the RBD and of those, 3 donors had been vaccinated. Most of the infected participants had either asymptomatic or mild outpatient COVID-19, and infections most likely represent Delta (Nov 2020-Nov 2021) and Omicron BA.1 (Dec 2021-present) variant waves [72].

Table 4. Demographics of pre-pandemic and ASU cohort donors. Median age was 25 y/o for the ASU cohort and most donors were male. No participants reported COVID-19-associated comorbidities.

		Pre-pandemic		ASU Cohort	
		n	%	n	%
Gender	Female	9	75	4	21.05
	Male	3	25	15	78.95
Age	19-25	0	0	8	42.11
	26-30	0	0	8	36.84
	31-35	1	8.3	1	5.26
	36-55	3	25	2	5.26
	56-70	6	50	0	0
	>70	2	16.7	0	0
Ethnicity	A/API	1	8.3	5	26.3
	W	11	91.7	8	42.1
	Hispanic	0	0	5	26.3
	N/A	0	0	1	5.3
Vaccine	mRNA-1273	-	-	6	31.6
	BNT162b2	-	-	7	36.8
	Ad26.COV2.S	-	-	3	15.8
	N/A	-	-	3	15.8
Sero-reactivity to RBD	Positive	5	42	15	78.95
	Negative	7	58	4	21.05

API: Asian Pacific Islander, A; Asian, W: White, His: Hispanic, RBD: Receptor Binding Domain. mRNA-1273: Moderna, BNT16b2: Pfizer, As26.COV2.S: Janssen/Johnson

To determine the dynamics of $CD8^+$ T cell recognition over time, we collected samples at 5 different time points: time of recruitment or baseline, 2 weeks, and at 1, 3, 6, and 12 months after the first sample collection. At the end of sample collection, 16/19 individuals were vaccinated with either mRNA-1273 (n=6), BNT162b2 (n=7) or ad26.COV2.S (n=3) vaccines and three individuals remained unvaccinated (2 dropped out). For each time

point, the days post vaccination/infection are detailed as well as the sero-reactivity (IgG) to RBD of SARS-CoV-2 spike and N protein of 7 different human coronaviruses (SARS-CoV-2, SARS-CoV, MERS-CoV, 229E-CoV, NL-63-CoV, OC-43 CoV and HKU-1 CoV). Finally, to characterize pre-existing responses to SAR-CoV-2 we tested pre-pandemic samples (n=12) collected between 2017 and early 2019, well before the start of the SARS-CoV-2 pandemic, declared on the 11th of March of 2020 [73].

2.2.2 *Immunogenicity of SARS-CoV-2 CD8+ T cell peptides predicted by EnsembleMHC*

We used a consensus algorithm, EnMHC to feed the entire sequence of the SARS-CoV-2 proteome and obtained a list of 658 high-confidence structural and non-structural peptides predicted to bind to 52 of the most common HLA-I alleles worldwide (Fig. 5). The current study prioritized the validation of all predicted structural peptides (n=108) to determine immunogenic hotspots that can be targeted with vaccines, especially if they are highly conserved among common cold coronaviruses (CCCoVs) and emergent variants of concern (VOC). Analysis of non-structural proteins is equally important given that proteins like nsp12 (RdRp) and nsp 3 (endonuclease) are essential for the completion of the viral cycle, are highly conserved among HCoVs and VOC, and are known to be expressed in high enough quantities to be considered immunoprevalent [74]. However, given the amount of PBMCs available for analysis, and that previous SARS-CoV studies show that structural proteins co-dominate the response, we decided to prioritize testing of structural proteins [75].

To determine peptide reactivity, ELISpot IFN-γ secretion was quantified as the number of spot-forming units (SFU) per 10^6 PBMCs (peripheral mononuclear cells) – see Methods 2.4.7. To be considered reactive or immunogenic, the number of SFU had to: 1) be above the 2-fold threshold (sample SFU/background SFU) and 2) be above the

background number of SFU plus 3 standard deviations. When used together, these two parameters reduce the false discovery rate (FDR) when less than 5 replicates per peptide are tested [76, 77].

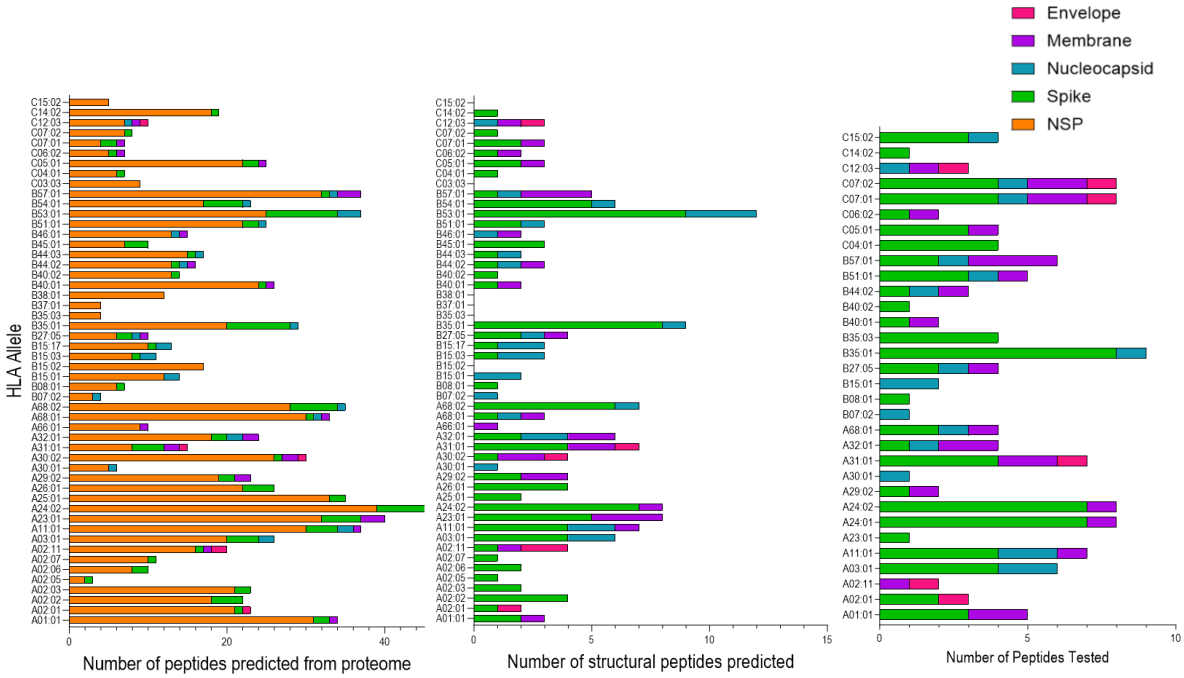


Figure 5. Distribution of EnsembleMHC predicted peptides from the SARS-CoV-2 proteome among 52 HLA-I alleles. A, Binding capacity to nsp and structural peptides predicted with EnMHC. B. Enhancement of structural peptides. C. Peptides tested according to cohort HLA restriction. NSP: non-structural. Adapted from (Wilson, et al. 2022).

Due to the HLA restriction of the pre-pandemic and ASU cohorts, we validated the immunogenicity of 93 out of the 108 structural peptides, restricted to 31 different HLA class I A, B, and C alleles. The group of tested peptides and HLA combinations covers>99.56% of the North American population [78] as expected from the diverse group of students that attend the university (Fig. 6). The median number of predicted peptides from the entire proteome of SARS-CoV-2 per allele is 17 with a range between 3 and 47 peptides per allele. HLA-A*23:01 and 24:02 have the highest binding capacity among all

class I alleles, while A*02:05, B*07:02, B35:03, B*37:01, and C*12:05 have the lowest. As reported previously [43], inter-allele binding capacities have higher variation among structural proteins than for the entire proteome, and this is exemplified by the higher number (n=6) of HLAs predicted not to bind any structural peptides after analysis with EnsembleMHC (e.g. HLA-B*15:02, B35:03). The binding capacity range for structural peptides is between 5 and 23 peptides and between 23 and 134 for non-structural proteins. As a group, HLA class I-A alleles are predicted to have the highest binding capacity with an average of 20.5 structural peptides and as many as 533 peptides for the entire proteome (Table 3, 5). After testing for reactivity, the observed binding capacity to structural epitopes was not significantly different from the predicted value (Wilcoxon-matched ranked pairs $P>0.05$) (Fig 7). The observed binding capacity may be higher given that not all the predicted peptides could be tested in every instance due to a limited number of available cells per sample. Of note, the observed binding capacity does not take into account promiscuity of HLA binding, peptide presentation by different MHC molecules present in the same donor or from HLAs that belong to the same supertype and that in theory could present the same peptide. In this respect, careful itemization was done per peptide to determine if a peptide was restricted to more than one HLA per donor. Of the 93 tested peptides, 17 were tested in individuals who were restricted to two HLAs that could present the same individual peptide. For instance, peptide KAYNVTQAF^{N} was tested in at least 5 different individuals but one donor was restricted to B*57:01 and C*06:02, both MHCs predicted to present this peptide. To directly determine the HLA restriction of the individual peptides, future experiments should include the identification of HLA-bound peptides by liquid chromatography-tandem mass spectrometry (LC)MS/MS) of peptide-pulsed monoallelic cell lines or by peptide titration or limiting dilution analysis [79, 80].

After testing reactivity with ELISpot IFN-γ secretion assays we found that 83% of tested peptides were immunogenic (above 2-fold threshold) (Fig. 8). This suggests, that EnMHC can predict immunogenic epitopes with high precision. When estimating immunogenicity using the IEDB platform [22] masking P2, P7, and P9 residues, 65/108 peptides have a positive score predictive of reactivity, or 56.5% of the 108 peptides were predicted to be immunogenic. This contrasts with the observed epitopes that we found to be immunogenic, 30 of which were missed by the IEDB algorithm. A literature review of the tested and immunogenic SARS-CoV-2 structural peptides revealed that of the 108 epitopes predicted by EnMHC, 65 of the 92 peptides validated in this study were also reported in previous studies. About 36/65 were individually validated by ICS (Intracellular Cytokine Staining), multimer staining, ELISpot, or AIM (Activation Induced marker) assays while the rest (29/65) were tested using peptide pools [81]. The current study individually validates the immunogenicity of those 29 SARS-CoV-2 $CD8^+$ T cell epitopes, plus 27 additional structural epitopes not reported elsewhere. In total, we show the individual immunogenicity of 56 peptides and confirm/expand the reported HLA restriction of 75 epitopes. We also provide evidence of the HLA restriction of 6 epitopes that were previously validated using a high throughput multiplexed assay designed to isolate TCRs after activation with peptide pools but not to define peptide HLA restriction (Appendix A) [82].

The predicted and observed personal binding capacities of donors do not differ significantly from each other, highlighting the accuracy of EnMHC and opening the possibility of using such platforms to map the epitope repertoire of individuals and from populations to determine their association, if any, to disease outcome. Specifically to SARS-CoV-2, the binding capacity to nsps is different than for structural peptides (Fig. 5, 7) possibly due to proteomic differences between them and further post-translation

modifications (PTMs) in structural proteins like Spike. Although the consensus algorithms in EnMHC don't consider PTMs, intrinsic characteristics of the peptides are reflected in the differences in binding capacity. Future studies that include data mining of HLA-typed cohorts associated with disease outcomes will clarify the importance of binding capacity as a criterion for evaluating the disease susceptibility of individuals or populations instead of individual HLA alleles associative studies. In our cohort, the median binding capacity per individual was around 95 peptides, both from non-structural and structural peptides, 84% of donors had a binding capacity higher than the median. The donor with the lowest binding capacity, ASC-95, is monoallelic for each of the HLA class I alleles, explaining the smaller peptide repertoire.

2.2.3. Immunodominance and Immunoprevalence of SARS-CoV-2 structural proteins

We screened $CD8^+$ T cell reactivity in vitro by stimulating peripheral blood mononuclear cells (PBMCs) for 10 days with individual peptides derived from SARS-CoV-2 structural proteins. T cell reactivity was tested by IFN-γ ELISpot secretion and Activation Induced Marker (AIM) assays. Results from a reactive and an unreactive sample are shown in figure 9. A total of 76 peptides were immunogenic (81.7%), and 17 (18.3%) were not reactive in either pre-pandemic or ASU cohort. T cells from the pre-pandemic cohort recognized 37.3% (n=22/59) of tested peptides while the student cohort, a mix of convalescent, vaccinated, and unvaccinated donors, recognized 86% of tested peptides at the time of recruitment (n=73/85). Median reactivity as measured by INF-γ secretion SFU (sport forming units) to structural peptides in pre-pandemic samples was about 15%. For comparison, 37.3% of SARS-CoV-2 tested peptides were recognized by pre-pandemic samples, or less than half of those that were recognized by the ASU cohort at the start of the study. Although the pre-pandemic and ASU cohorts are unrelated, with

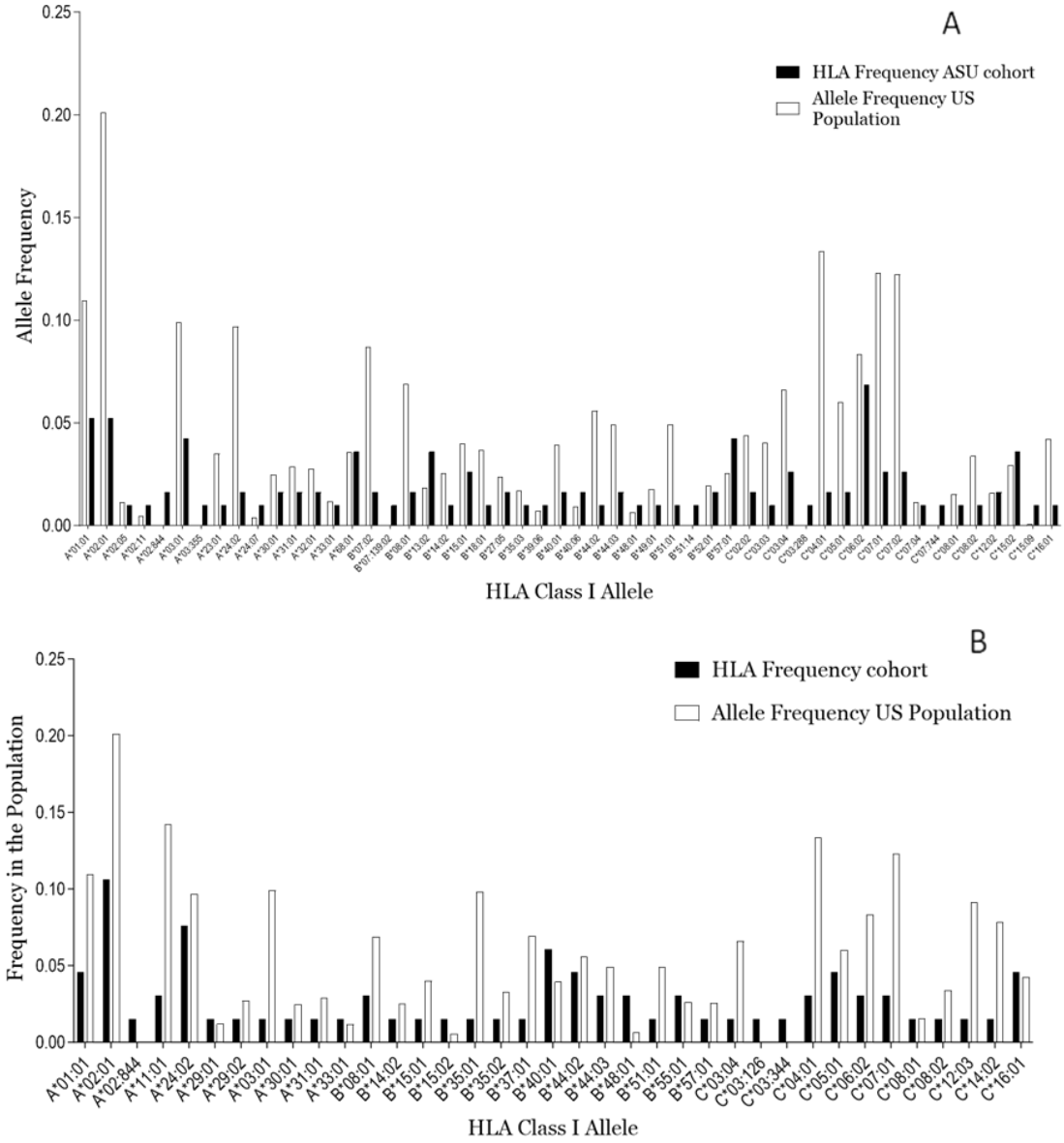


Figure 6. (**A**) Frequency of common HLA alleles in the ASU (n=19) and (**B**) pre-pandemic (n=12) cohort compared to the frequencies in the US population (n=3500). Allele frequencies obtained from allelefrequencies.net [83].

different donors and HLA restrictions, the median reactivity was lower for the pre-pandemic samples (about 15%) than for ASU cohort samples at the time of recruitment (61%). The percentage of epitopes recognized, and the magnitude of responses was higher after 3- and 6-months post-recruitment, when most donors had been infected, vaccinated,

or both. Background reactivity was 2 to 7 times higher at the end of the study, 12 months post recruitment.

Each donor recognized an average of 10.2 epitopes with a median of 9.5 and a range between 2 and 21 epitopes. The donor with the least recognized epitopes (n=2) was restricted to HLAs A*02:01, B*52:01, and C*12:02, and was also considered monoallelic for A, B, and C alleles (NULL). As a result, the predicted binding capacity consisted of 5 structural peptides, of which one was from the envelope and the other 4 from the spike protein. None were immunogenic, until after vaccination. The donors that recognized the greatest number of peptides were mostly restricted to HLA-A*01:01, A*03:01, -B*57:01, and -C*06:02. Although we could not test the entire number of predicted peptides per

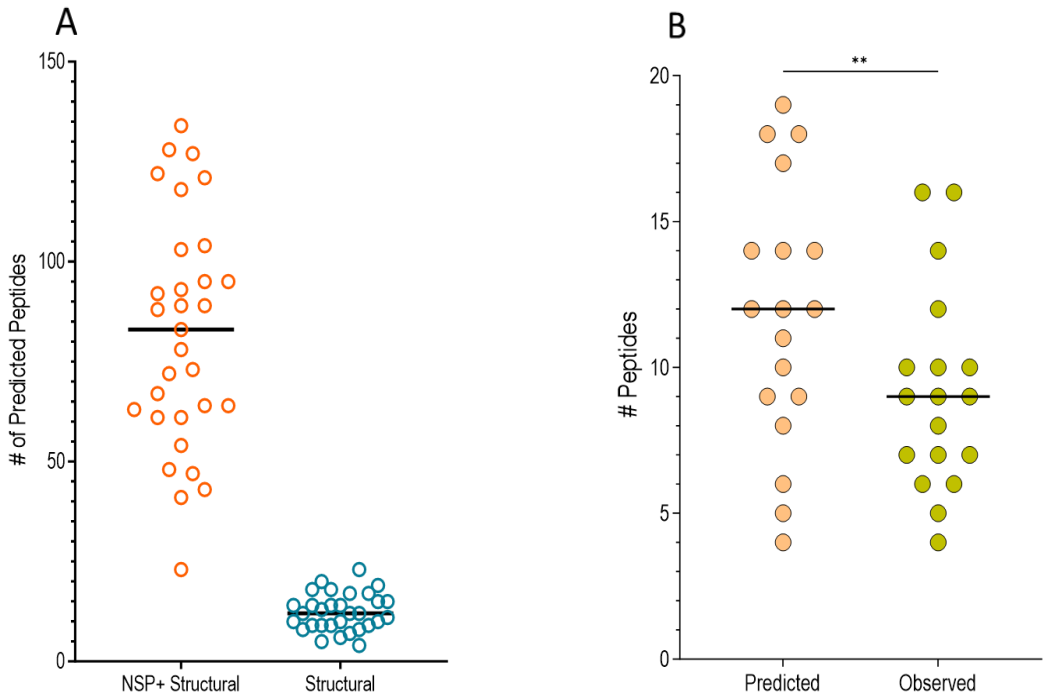


Figure 7. A. Comparison of the binding capacity per individual (n=31) to non-structural and structural proteins. **B.** The observed binding capacity of tested ASU donors (n=19) is significantly similar than the predicted binding capacity (**P=0.1226 Wilcoxon paired test of similarity).

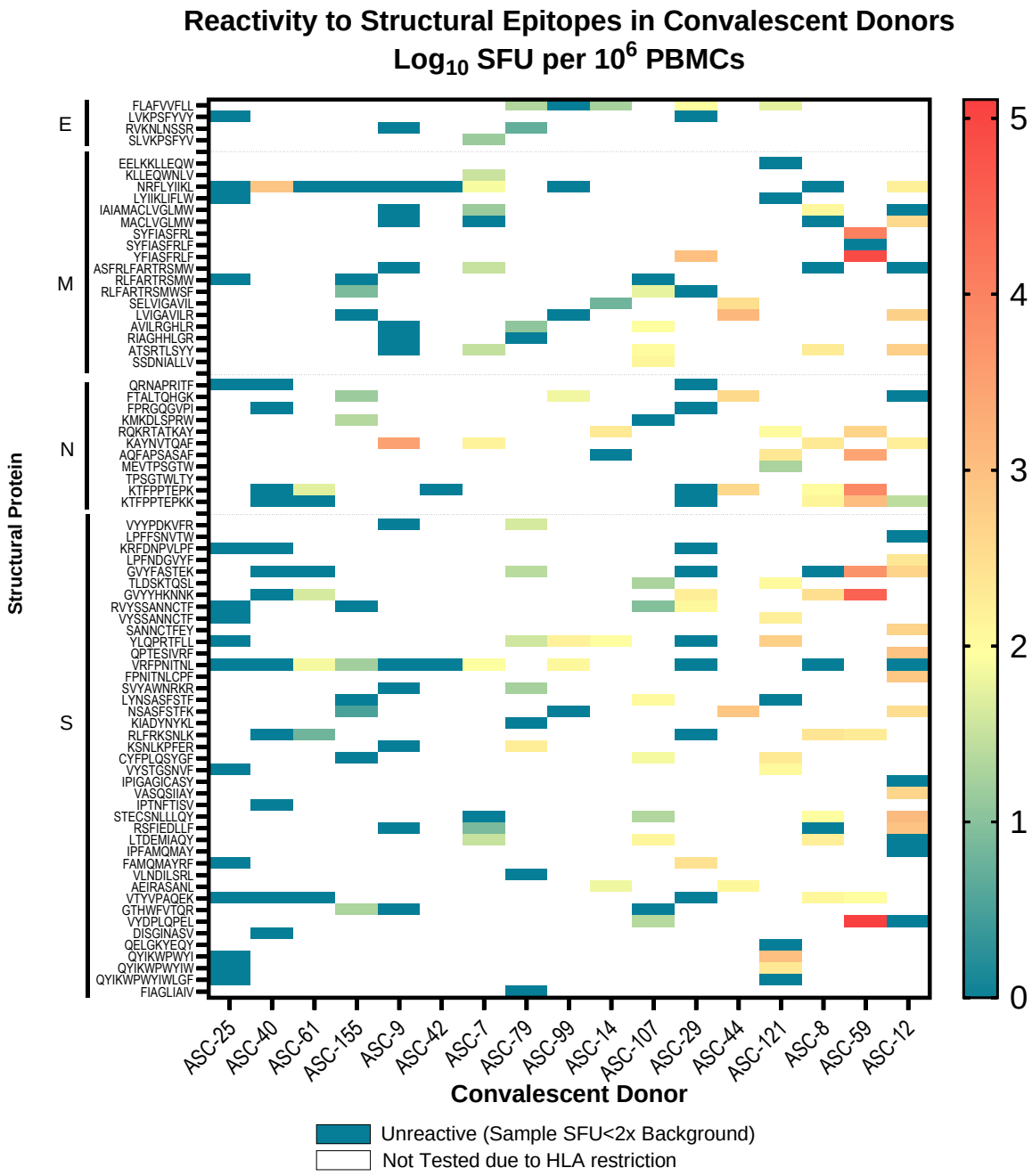


Figure 8. Heat map depicting the magnitude of response to SARS-CoV-2 CD8⁺ T cell epitopes at the time of recruitment into the ASU cohort study. Epitopes are organized by rows according to position in the origin protein, from N- to C-terminal. Each column represents a different donor. Reactive epitopes have a reactivity over 2 fold over the background. E: Envelope, M: Membrane, N: Nucleocapsid, S: Spike.

Table 5. Binding capacity per HLA-I Allele. Binding capacity is higher for nsps and those HLA-I alleles that are most common in the global population. Note the low binding capacity of HLA-B*35:01, B*37:01 and C*12:02 in contrast to those of A*01:01, A*11:01, A*68:01, B*57:01 and C*05:01. NSP: non-structural proteins.

HLA Allele	NSP	Spike	Nucleocapsid	Membrane	Envelope
A01:01	31	2		1	
A02:01	21	1			1
A02:02	18	4			
A02:03	21	2			
A02:05	2	1			
A02:06	8	2			
A02:07	10	1			
A02:11	16	1		1	2
A03:01	20	4	2		
A11:01	30	4	2	1	
A23:01	32	5		3	
A24:02	39	7		1	
A25:01	33	2			
A26:01	22	4			
A29:02	19	2		2	
A30:01	5		1		
A30:02	26	1		2	1
A31:01	8	4		2	1
A32:01	18	2	2	2	
A66:01	9			1	
A68:01	30	1	1	1	
A68:02	28	6	1		

HLA Allele	NSP	Spike	Nucleocapsid	Membrane	Envelope
B07:02	3		1		
B08:01	6	1			
B15:01	12		2		
B15:02	17				
B15:03	8	1	2		
B15:17	10	1	2		
B27:05	6	2	1	1	
B35:01	20	8	1		
B35:03	4				
B37:01	4				
B38:01	12				
B40:01	24	1		1	
B40:02	13	1			
B44:02	13	1	1	1	
B44:03	15	1	1		
B45:01	7	3			
B46:01	13		1	1	
B51:01	22	2	1		
B53:01	25	9	3		
B54:01	17	5	1		
B57:01	32	1	1	3	

HLA Allele	NSP	Spike	Nucleocapsid	Membrane	Envelope
C03:03	9				
C04:01	6	1			
C05:01	22	2		1	
C06:02	5	1		1	
C07:01	4	2		1	
C07:02	7	1			
C12:03	7		1	1	1
C14:02	18	1			
C15:02	5				

genotype (donor) due to a limited number of cells per sample, certain donors consistently recognized between 65-100% of tested peptides (ASC-12, ASC-59, ASC-8, ASC-7). Overall, donor samples became increasingly reactive over time, background reactivity was higher at later time points. Time points 3 and 6 months, the percent of peptides recognized increased from 61% to >75% when all participants were either convalescent, vaccinated, or both. Of note, from the samples at the time of recruitment (61% reactivity), 73% were either convalescent, vaccinated, or both, and 17% were self-reported as uninfected and unvaccinated. Since sample collection started 1 year after the start of the pandemic and about 9 months after the virus had spread throughout North America, it is possible that

self-reported uninfected individuals were either asymptomatic during infections or had been exposed to someone with the virus.

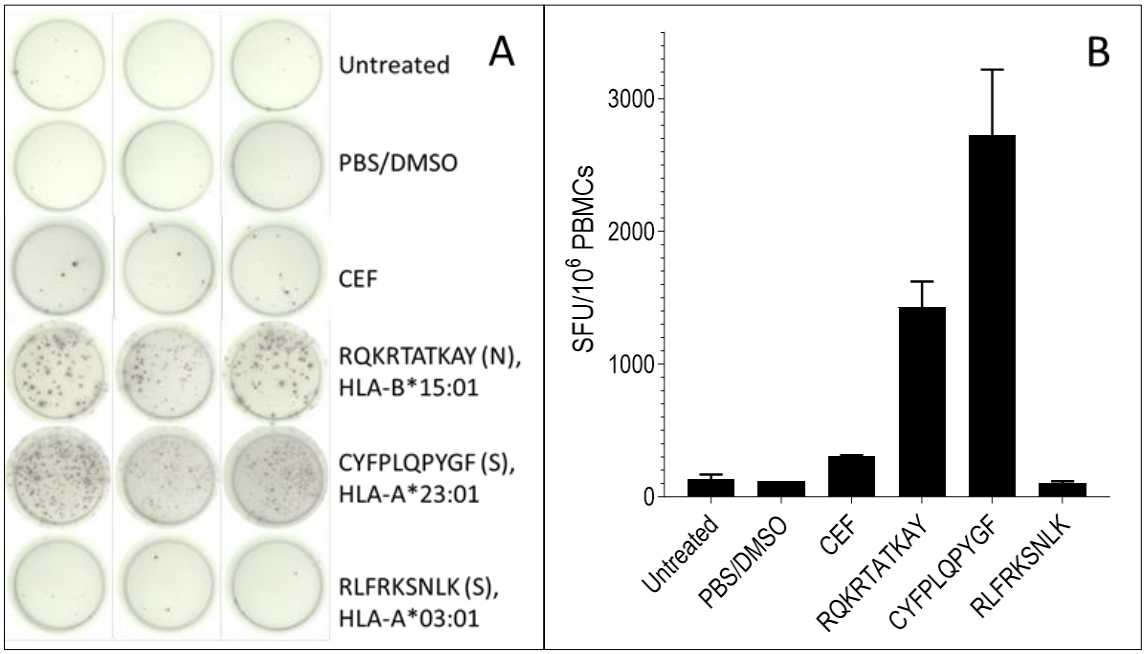


Figure 9. CD8+ T cell reactivity to predicted peptides measured by IFN-γ secretion ELISpot. A. Representative image after 10-day stimulation with single peptides of reactive and unreactive peptides. Each peptide or condition was tested in triplicate wells. Each spot represents an IFN-γ secreting cell unit counted as a spot-forming unit. B. Histogram of the average number of SFU per 10^6 PBMCs with the standard deviation. Convalescent donor ASC-116, sero-reactive to RBD, 1-month post-vaccination. Negative controls: Untreated and PBS/DMSO wells. Positive control: CEF pool.

The number of positive responses is proportional to the size of the protein (Spearman correlation r=0.8, P=0.33, Figure 10), this is supported by previous reports that show a direct correlation between the number of recognized targets and the size of the protein with proportional distribution of epitopes along the SARS-CoV-2 genome [84]. Because the selection of epitopes with EnMHC was restrictive, we did not test epitopes alongside the entire sequence of each structural protein to differentiate hotspots of immunogenicity. However, we can see that the number of immunogenic targets increases with the size of

the protein. Much of the response (50%) was represented by spike epitopes, followed by membrane (23%), then nucleocapsid (11.5%), and Envelope (3.5%) epitopes.

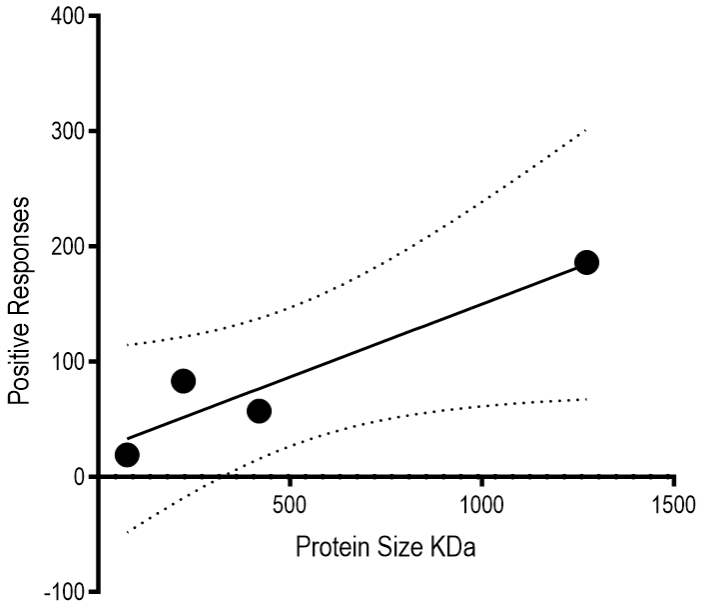


Figure 10. Spearman correlation of positive responses per SARS-CoV-2 structural protein. (r^2=0.896). Tested epitopes were evenly distributed across the different proteins in proportion to size.

Overall, 30% (26/85) of the peptides tested on the ASU cohort samples, were recognized by 2 or more donors and were consistently recognized at each time point. We found that a total of 15 epitopes dominated the response and were distributed evenly among the structural proteins S, M, and N even after vaccination (Table 6). Epitopes from spike (n=7), membrane (n=4), and nucleocapsid (n=4) were highly reactive through all time points and were recognized by 3 or more donors. Of note, FLAFVVFLL was the only epitope from envelope that was highly reactive (up to 14-fold over background) and was recognized consistently between 3-5 different donors at each time point and was reactive in one pre-pandemic donor.

2.2.4 Conservation of Immunodominant Epitopes Across VOC

To assess the impact of mutations on the recognition of $CD8^+$ T cell predicted peptides from structural proteins, we tested reactivity by ELISpot INF-γ secretion assays of individual peptides derived from the ancestral sequence and whenever mutated, the variant epitope (Table 7). The ancestral sequence was aligned against Alpha (B.1.1.7), Beta (B.1.137), Delta (B.1.617.2), Gamma, and Omicron (B.1.1.529 BA.4 and BA.5) VOC (Appendix B). We found that from the 108 structural predicted peptides, 17 peptides had at least 1 amino acid change in one or more VOC. Most of the mutated peptides were from the spike protein (n=15) while 2 mutations were found in membrane and 1 in nucleocapsid proteins. In some peptides, mutations were found across multiple VOCs, for a total of 23 mutations of which 20 occurred in the spike protein. Most mutations (15/23) were associated with the Omicron subvariants and in one case ($VLNDILSRL^{WT}$), the mutation characteristic of the subvariant BA.1 (VLNDI**F**SRL-changed residue is in bold letter) reverted to the ancestral sequence in the new subvariants BA.4 and BA.5. Due to HLA restriction and/or limited number of PBMCs, we tested 10/17 variant peptides and found that 9 were immunogenic in at least one donor. In some cases, the variant epitope induced a higher reactivity than the ancestral counterpart as in the spike epitope $GVYFASTEK^{WT}$, with a change at anchor residue P7 from threonine to isoleucine ($GVYFAS\mathbf{I}EK^{O}$); the magnitude of response was up to 4-fold that of the ancestral version. Variant epitopes were 89% conserved (one amino acid change), for example, $IAIAMACLVGLMW^{WT}$ (M) had a substitution at the third amino acid (I82T), and KLLEQWNLV (M) was conserved across all variants except for Omicron. A Q19E substitution in all three subvariants did not affect reactivity that was >2-fold over background when tested in HLA-A*02:11 and -A*02:01 restricted donors.

In one instance, the variant (VLNDI**F**SRL) was immunogenic while the ancestral peptide (VLNDILSRL) was not. One RBD (receptor binding domain) epitope LYNSASFSTFK^{WT} (S), had 3 and 4 amino acid substitutions from the BA.1 and BA.4/5 Omicron subvariants respectively (Table 7, Figure 11). To end, alignment of the immunodominant epitopes with VOC shows high conservation, 87% (13/15) have 100% sequence identity. The other two epitopes, GVYYHKNNK^{WT} and NSASFSTFK^{WT} from the NTD and the RBD respectively, had changes related to the Omicron variant. The $_{142}$GVYYHKNNK$_{150}^{WT}$ peptide had a G to D substitution at the P1 residue and the $_{370}$NSASFSTFK$_{378}^{WT}$ had three amino acid changes (N**L**A**P**F**F**TFK and N**F**A**P**F**FA**FK) that did not decrease the magnitude of response (Table 7).

Table 6. Immunodominant CD8$^+$ T cell structural epitopes.

PROTEIN	Peptide Sequence	Position		Reactive Donors	HLA Restriction		MHC Flurry			Variants
		Start	End		EnMHC	Observed	BA	Processing	Presentation	Alpha
E	FLAFVVFLL	20	28	3	A*02:01, A*02:11	A*02:01	11.566	0.013	0.880	Not mutated
						B*08:01	110.193	0.013	0.451	
M	LVIGAVILR	138	146	3	A*66:01,A*68:01	A*68:01	27.958	0.360	0.914	Not mutated
M	MACLVGLMW	84	92	3	B57:01	B*57:01	34.998	0.247	0.851	Not mutated
M	NRFLYIIKL	43	51	6	B27:05|C06:02|C07:01	C*06:02	137.080	*		
						C*07:01	37.500	0.957	0.985	
N	FTALTQHGK	50	58	3	A68:01	A*68:01	24.872	0.101	0.826	Not mutated
N	KAYNVTQAF	263	271	3	A32:01|B15:03|B15:17 B46:01|B57:01|C12:03	B*15:01	20.330	*		Not mutated
						B*57:01	41.484	0.754	0.967	
						C*12:03	29.902	0.754	0.976	
N	KTFPPTEPK	358	366	3	A03:01|A11:01|A30:01	A*03:01	26.984	0.254	0.883	Not mutated
						A*02:01	24843	*		
						A30:01	21.295	0.484	0.955	
N	RQKRTATKAY	256	265	3	B15:01	B*15:01	43.338	*		Not mutated
S	GVYYHKNNK	142	150	3	A03:01	A*03:01	27.860	*		Mutated Omicron
S	LTDEMIAQY	860	868	3	A01:01	A*01:01	25.381	0.845	0.985	Not mutated
S	NSASFSTFK	370	378	2	A68:01	A*68:01	24.119	0.341	0.920	Mutated Omicron
S	RSFIEDLLF	810	818	3	B15:17|B57:01	B57:01	45.422	0.575	0.934	Not mutated
S	STECSNLLLQY	741	751	4	A01:01	A*01:01	34.384	0.380	0.903	Not mutated
S	VRFPNITNL	322	330	5	B27:05|C06:02|C07:01 C07:02	C*06:02	28.967	0.813	0.981	Not mutated
						C*07:01	56.630	*		
						C*07:02	95.920	*		
S	YLQPRTFLL	264	272	4	A02:01|A02:02|A02:03|A02:05	A*02:01	4.300	*		Not mutated
						A*02:11	12.338	0.670	0.986	
					|A02:11|B08:01	B*08:02	44.736	0.670	0.953	
						C*05:01	79.698	0.670	0.920	

Position, source protein, and HLA predicted restriction. Observed HLA restriction and BA (IC50), processing, and presentation scores obtained with MHC Flurry. S: Spike, N: Nucleocapsid, M: Membrane, E: Envelope. BA: Binding affinity in IC50; (*): Epitopes not predicted to bind the EnMHC predicted HLAs under MHC Flurry, BA was determined using NetHCpan 4.0.

Table 7. Peptide sequences from ancestral and VOC structural SARS-CoV-2 peptides.

Protein	Start	End	Ancestral Sequence	Reactive	Sequence	EnMHC HLA	Variant	Reactive	Location
M	11	21	EELKKLLEQW	Yes*	EELKKLLE**E**W	B44:02	Omicron	Not Tested	NTD-M
M	15	23	KLLEQWNLV	Yes	KLLE**E**WNLV	A02:11	Omicron	Yes	NTD-M
N	9	17	QRNAPRITF	No	QRNA**L**RITF	B27:05	Omicron	Not Tested	NTD-N
S	8	16	LPLVSSQCV	No*	LPLVS**I**QCV	B51:01	B.1.429	Not Tested	SP/NTD
S	72	81	KRFDNPVLPF	Yes	KRF**G**NPVLPF	B27:05	B.1.526	Not Tested	NTD
S	84	92	GVYFASTEK	Yes	GVYFASIEK	A03:01|A11:01	B.1.526 Omicron	Yes	NTD
S	104	112	CPFGEVFNA	Not Tested	CPF**D**EVFNA	B54:01	Omicron	Not Tested	NTD
S	142	150	GVYYHKNNK	Yes	**D**VYYHKNNK	A03:01	Omicron	Yes	NTD
S	150	158	SEFRVYSSA	Not Tested	SES**R**VYSSA	B45:01	B.1.526	Not Tested	NTD
S	368	377	LYNSASFSTF	Yes	LYN**LAPFFTF** LYN**FAPFFAF**	A23:01|A24:02	BA.1 BA.4/5	Not Tested	RBD
S	370	378	NSASFSTFK	Yes	N**L**APFFTFK N**F**APFF**A**FK	A68:01	BA.1 BA.4/5	Yes Yes	RBD
S	417	425	KIADYNYKL	No*	**N**IADYNYKL	A02:02|A02:06|A32:01	Omicron	Yes	RBD
S	493	492	CYFPLQSYGF	Yes	CYFPLQ**P**YGF -YFPLQ**P**YGF CYFPL**R**SYSF	A23:01|A24:02	B.1.1.7 B.1.1.7 BA.4/5	Yes No Yes	RBD
S	709	717	IPTNFTISV	No	IP**I**NFTISV	B15:03	B.1.1.7	Not Tested	S2
S	971	979	VLNDILSRL	No	VLNDIL**A**RL VLNDI**F**SRL VLNDI**L**SRL	A02:02|A02:03	B.1.1.7 BA.1 A-BA.4/5	Not Tested Yes No	HR1
S	1060	1069	VTYVPAQEK	Yes	VTYVPA**H**EK	A03:01|A11:01	B.1.617.1	Yes	HR1
S	1163	1171	DISGINASV	Not Tested	DISGINAS**F**	A68:02	P2	Not Tested	HR2

M: Membrane, N: Nucleocapsid; S: Spike. VOC: Alpha (B.1.1.7), Beta (B.1.351), Delta (B.1.617), Iota (B.1.526), P.1 (Gamma), Omicron (BA.1, BA.4, BA.5). (*): only tested once. The bold letter represents the changed amino acid. A: ancestral.

2.2.5 Vaccination, Infection, and Reactivity to CD8⁺ T Cell Epitopes Over Time

Reactivity to SARS-CoV-2 epitopes was significantly higher at the time of recruitment than for pre-pandemic samples (Kruskal-Wallis and Dunn's multiple comparison test $P<0.0079$). However, at the time of recruitment less than half of participants (9/19) self-reported to have ever been infected with SARS-CoV-2. Of those, 4 were unvaccinated and 5 had completed the full vaccination schedule that consists of 2 doses of either Moderna or Pfizer mRNA platforms or one dose of the J&J vaccine. The rest of the participants, 10/19 donors had asymptomatic or mild COVID-19 prior to the start of the study (2-6 weeks prior enrollment) and of those, 3 had been vaccinated (Table 8). Given the differences in exposure to the virus, convalescence, and vaccination, samples were reorganized along the 5 time points according to infection status, and time of vaccination. Participant samples were organized according to time of infection (if unvaccinated), or

time of vaccinated (uninfected or convalescent). For most donors (15/19) who were convalescent or vaccinated at the start of the study, there was no baseline measurement, instead, their first sample was placed according to the number of weeks or months after infection or vaccination. For example, a few recruitment samples were placed at the 1-month time point because they had completed the vaccination schedule, 2 doses of mRNA or 1 dose of Ad26.COV2.S vaccine.

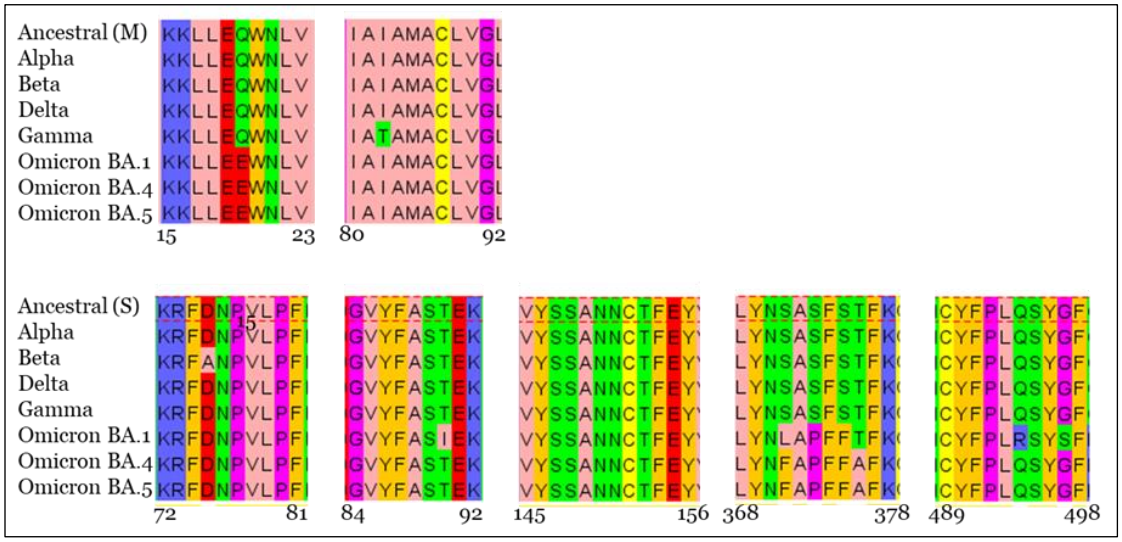


Figure 11. Alignment of variant epitopes from the Membrane (Top) and Spike (bottom) protein. Zappo color scheme showing physicochemical properties at each amino acid position. Most variant epitopes had only one amino change, except for LYNSASFSTF^{WT} which had 4 amino acid changes mostly associated with the Omicron subvariants. Pink: Aliphatic/hydrophobic (ILVAM); Orange: Aromatic (FWY); Blue: Positive (KRH); Red: Negative (DE); Green: Hydrophilic (STNQ); Fuscia: conformationally special (PG); Yellow: Cysteine (85).

Table 8. Fraction of epitopes recognized by time point.

Time Point	Donor	Status	IgG for RBD	Vaccination Status	Fraction Recognized Epitopes	Median Fraction	Total # Peptides Recognized	Total # Peptides Tested	Median Reactivity	Total Donor
R	1	Uninfected	Negative	UnVx	0.650					
	2	Uninfected	Negative	UnVx	0.000					
	3	Uninfected	Negative	Vx	0.000					
	4	Uninfected	Negative	UnVx	0.833	0.325	23	60	0.722	4
2W	5	Convalescent	Positive	UnVx	0.625					
	6	Convalescent	Positive	UnVx	0.214					
	7	Convalescent	Positive	UnVx	0.833					
	8	Convalescent	Positive	UnVx	0.083					
	9	Convalescent	Positive	UnVx	1.000					
	10	Convalescent	Positive	Vx	0.500					
	11	Convalescent	Positive	UnVx	0.917					
	12	Uninfected	Negative	Vx	0.222					
	13	Uninfected	Positive	Vx	0.583					
	14	Convalescent	Positive	Vx	0.778					
	15	Uninfected	Positive	Vx	0.778	0.625	67	121	2.299	11
1M	16	Convalescent	Positive	Vx	0.524					
	17	Convalescent	Positive	UnVx	0.273					
	18	Uninfected	Positive	Vx	0.000					
	19	Convalescent	Positive	Vx	0.700					
	20	Convalescent	Positive	Vx	0.333					
	21	Convalescent	Positive	Vx	0.200					
	22	Uninfected	Positive	Vx	0.182					
	23	Uninfected	Positive	Vx	0.000					
	24	Convalescent	Positive	Vx	0.818					
	25	Uninfected	Positive	Vx	0.625					
	26	Uninfected	Positive	Vx	0.600					
	27	Convalescent	Positive	Vx	0.800	0.42857	58	118	1.654	12
3M	28	Convalescent	Positive	Vx	0.818					
	29	Convalescent	Positive	UnVx	0.353					
	30	Uninfected	Positive	Vx	0.667					
	31	Convalescent	Positive	Vx	0.250					
	32	Convalescent	Positive	Vx	0.375					
	33	Convalescent	Positive	Vx	0.125					
	34	Uninfected	Positive	Vx	1.000					
	35	Uninfected	Positive	Vx	0.636					
	36	Uninfected	Positive	Vx	0.625					
	37	Uninfected	Positive	Vx	0.000					
	38	Uninfected	Positive	Vx	0.909					
	39	Uninfected	Positive	Vx	0.286					
	40	Uninfected	Positive	Vx	0.000	0.375	65	113	3.521	13
6 M	41	Uninfected	Negative	Vx	1.000					
	42	Uninfected	Positive	Vx	0.000					
	43	Uninfected	Positive	Vx	0.429					
	44	Convalescent	Positive	Vx	0.286					
	45	Uninfected	Positive	Vx	0.667					
	46	Uninfected	Negative	Vx	0.875					
	47	Uninfected	Positive	Vx	0.300					
	48	Convalescent	Positive	Vx	1.000					
	49	Convalescent	Positive	Vx	1.000					
	50	Convalescent	Positive	Vx	0.000					
	51	Convalescent	Positive	UnVx	0.846					
	52	Uninfected	Positive	Vx	0.889					
	53	Uninfected	N/A	Vx	1.000	0.84615	70	98	3	13
12 M	54	Convalescent	Positive	Vx	1.000					
	55	Uninfected	Positive	Vx	0.750					
	56	Uninfected	Positive	Vx	0.800					
	57	Convalescent	Positive	Vx	0.250					
	58	Convalescent	Positive	Vx	0.250					
	59	Uninfected	Negative	Vx	0.692	0.72115	31	50	2.492	6

Donor ID was provided at the time of sample donation, not recruitment. To decrease the changes of identification, samples are not identified by the donor at the time of recruitment. All time points are arranged according to the approximate time of infection or vaccination. R: Recruitment; 2W: 2 weeks; 1M: 1 month; 3M: 3 months; 6M" 6 months; 12 M: 12 months. UnVx: Unvaccinated; Vx: Vaccinated. Median reactivity is fold change: SFU/10^6 PBMCs sample over the background.

Table 9. Vaccination, infection, and reactivity to human CoV status of ASU cohort at the time of recruitment. N: Nucleocapsid; R: Reactive.

	Access IgG		RAPID-ELISA								
Donor	Seroreactivity to RBD	Seroreactivity to N	SARS-CoV-2	OC-43	HKU-1	NL63	229E	SARS-CoV	MERS	Infection	Vaccination
1	*Negative*	*Negative*	*Negative*	R	R	R	R	R	*Negative*	No	No
2	R	R	R	R	R	R	R	R	*Negative*	No	Yes
3	*Negative*	*Negative*	R	R	R	R	R	*Negative*	*Negative*	No	Yes
4	R	R	R	R	R	R	R	*Negative*	*Negative*	Yes, <1 Month	No
5	R	R	*Negative*	R	R	R	R	R	*Negative*	Yes, <1 Month	No
6	*Negative*	*Negative*	R	R	R	R	R	*Negative*	*Negative*	No	No
7	R	*Negative*	*Negative*	R	R	R	R	R	*Negative*	Yes, <1 Month	No
8	*Negative*	*Negative*	*Negative*	*Negative*	NR	R	R	*Negative*	*Negative*	No	No
9	R	*Negative*	R	R	R	R	R	R	*Negative*	Yes, <1 Month	No
10	R	R	R	*Negative*	R	R	R	R	*Negative*	Yes, <1 Month	No
11	R	R	R	R	R	R	R	R	*Negative*	Yes, <3Months	No
12	R	*Negative*	R	R	R	R	R	R	*Negative*	Yes, <1 Month	No
13	R	R	*Negative*	R	R	R	R	R	R	Yes, <3Months	Yes
14	R	R	*Negative*	R	R	R	R	R	*Negative*	Yes, <3Months	Yes
15	R	R	*Negative*	R	*Negative*	R	R	R	*Negative*	Yes, <3Months	Yes
16	R	*Negative*	R	R	R	R	R	*Negative*	*Negative*	No	Yes
17	R	R	*Negative*	R	R	R	R	R	*Negative*	No	Yes
18	R	R	*Negative*	*Negative*	R	R	R	R	*Negative*	No	Yes
19	R	*Negative*	*Negative*	R	R	R	R	*Negative*	*Negative*	No	Yes

The recruitment time point (R) is now composed of 4 samples from uninfected unvaccinated participants, 11 samples were at the 2 weeks (2W) post-vaccination or infection time point, 13 samples at 1-2 months (1M), 12 samples at 3 moths (3M), 13 between 6 and 9 months (>6M) and 7 samples at 12 months (>12M) (Fig.12, Table 8, 9). To analyze the changes in $CD8^+$ T cell epitope recognition and magnitude of response over time, ELISpot INF-γ secretion to individual peptides was plotted as the number of SFU (spot forming units) per 1 million PBMCs over the different time points (Fig. 13, Appendix C). Reactivity or median reactivity was reported as the fold change of observed SFU per 10^6 PBMCs over the background, either untreated or PBS/DMSO whichever is lowest. For each time point, donor samples were grouped according to vaccination and convalescence status into 4 categories, uninfected unvaccinated (UU), uninfected vaccinated (UV), convalescent unvaccinated (CU) and convalescent vaccinated (CV). Participation throughout the year from the 19 donors was not consistent which led to some time points having only one donor as in the case of the 1M time point when only 1 individual was available for each of the UV and CU groups. Of note, the IFN-γ signal from the background

(untreated or PBS/DMSO) is 2-7 times higher at the 6- and 12-month time points which decreased the signal magnitude of the tested peptides (that were as reactive) and led to the elimination of a few samples.

Reactivity over time increases by the number of epitopes recognized and the magnitude of response. Median reactivity for every time point is significantly higher than pre-pandemic samples after the 2W time point and is higher than at any other time point at 3 months post-vaccination/infection. The percentage of recognized epitopes at the time of recruitment is 15% for pre-pandemic samples and starts to increase after infection or vaccination (Fig.14). At two weeks post-infection/vaccination the number of epitopes recognized by donor doubled and increased further after 1 month when all donors had finished the complete regiment of vaccines. By 12 months, the median percentage of recognized epitopes is about 72%. Due to the uncertainty of the date of infection, and mostly to the low number of convalescent-only individuals, it was not possible to discern a difference if any, in the magnitude and number of recognized epitopes between vaccinated-convalescent or convalescent-only individuals. At the 6-month time point, only one individual remained unvaccinated, the other 2 dropped out of the study after the 2W time point. In the end, the ASU donors for the 6M and 12M time points were either vaccinated with no reported infection or convalescent and vaccinated. The longest time interval after vaccination recorded for a donor was 16 months and 75 days post booster. Unfortunately, this donor was monoallelic for all three HLA class I alleles, and as a result, only 2 peptides could be tested for reactivity, from the 108 predicted by EnMHC. This donor was restricted to HLA-A*02:01, B*52:01, and C*12:02, unreactive to the two peptides from the E (FLAFVVFLL) and S (YLQPRTFLL) proteins at the earliest time points equivalent to 8 days post-vaccination completion with Moderna (mRNA-1273), 2 weeks post-vaccination (wpv) and 1.5 months post-vaccination (mpv). Peptides became

reactive at the 3-month blood collection, which is equivalent to ~4 mpv. Epitope recognition persisted during the 12-month follow-up that is 16 mpv.

Reactivity to Spike epitopes increased after 2 weeks post-infection or vaccination and remained elevated above the 2-fold threshold over background after 3 months. Median reactivity was highest at 3 months in UV donors which was 5-fold over background (Fig. 15) and was sustained above the 2-fold threshold 12-15 months after vaccination. $CD8^+$T cell reactivity was increased or maintained to epitopes from E, M, N, and Spike regardless of vaccination status after 6 months (Fig 16-18).

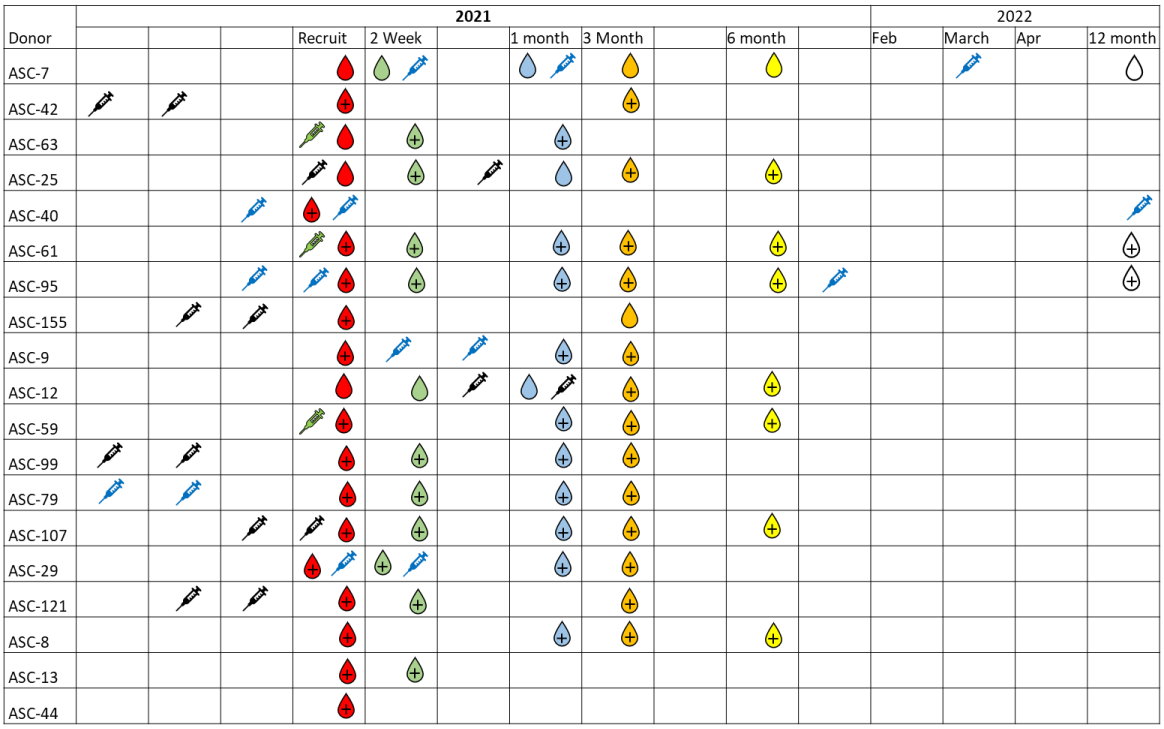


Figure 12. Timeline of sample collection and vaccination for ASU cohort. ASC identification number was the second number assigned after recruitment to maintain anonymity from experimenters. Drop colors: Red-Recruitment; Green-1 2 Weeks; Blue 1 Month, Orange- 3 Months, Yellow-6 Months; White-12 Months. White syringe: Pfizer; Blue: Moderna; Green: Janssen. (+): Sero-reactive for RBD of spike.

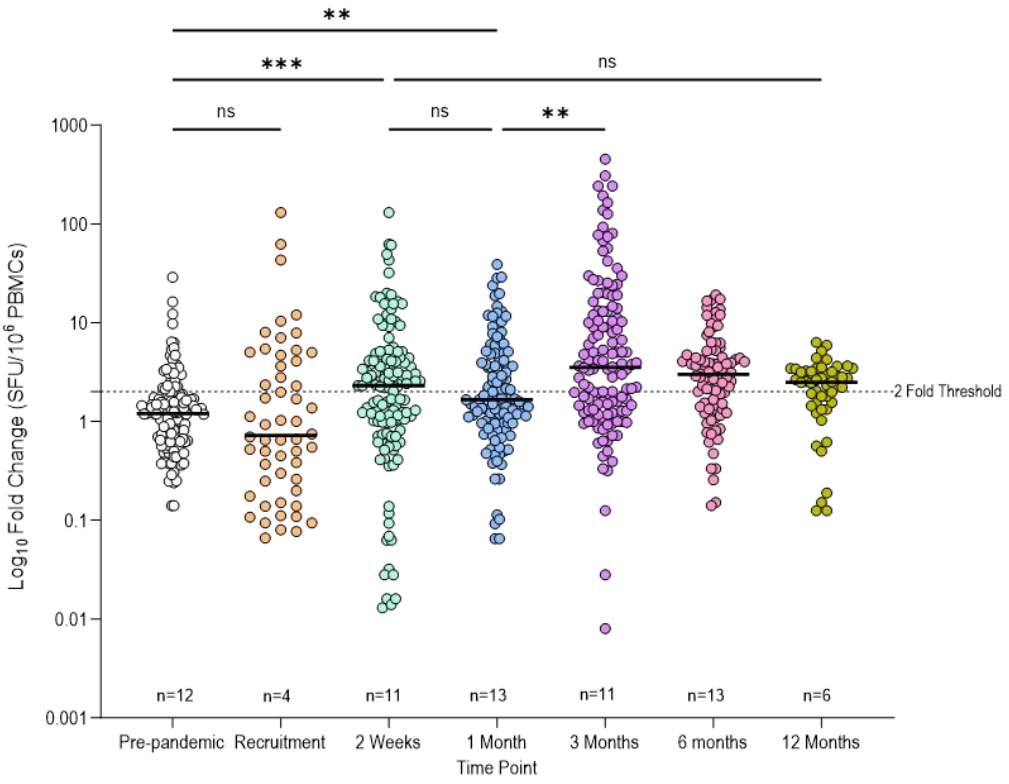


Figure 13. CD8+ T cell Reactivity to SARS-CoV-2 structural peptides tested by ELISpot IFN-γ secretion assay. The reactivity of all tested peptides per individual is graphed as the fold change (observed value over background) by the time point of sample collection. Median reactivity was significantly different from pre-pandemic samples except for samples collected at 1 month. Baseline samples are a mix of convalescent, vaccinated, or unvaccinated. By 6 months all individuals are either convalescent or vaccinated and convalescent. "n" represents the number of individuals at each time point. Individual dots represent the reactivity of each of the peptides tested.

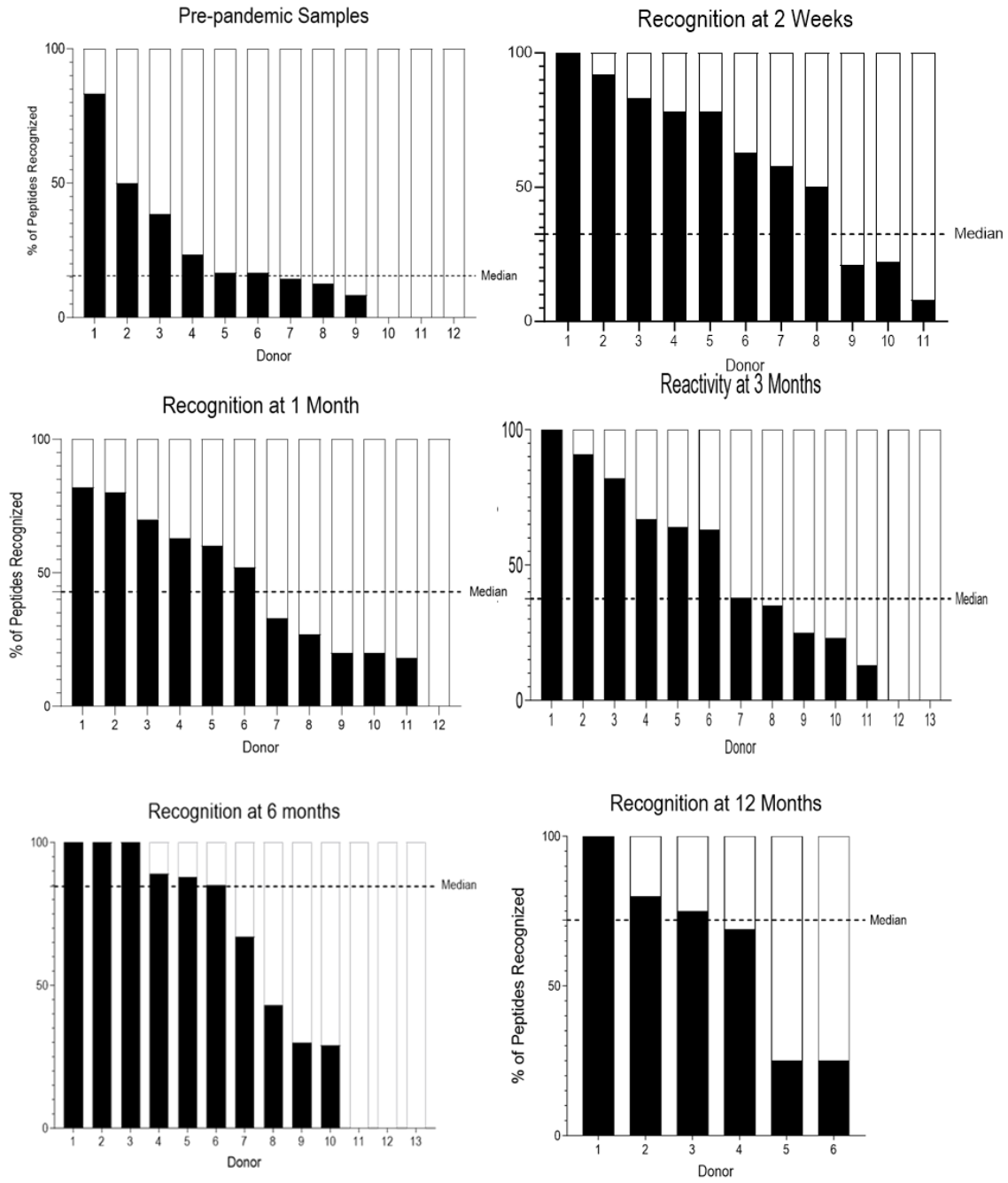


Figure 14. Epitope recognition after infection and/or vaccination. The number of epitopes recognized, and the magnitude of response is highest at 6 months.

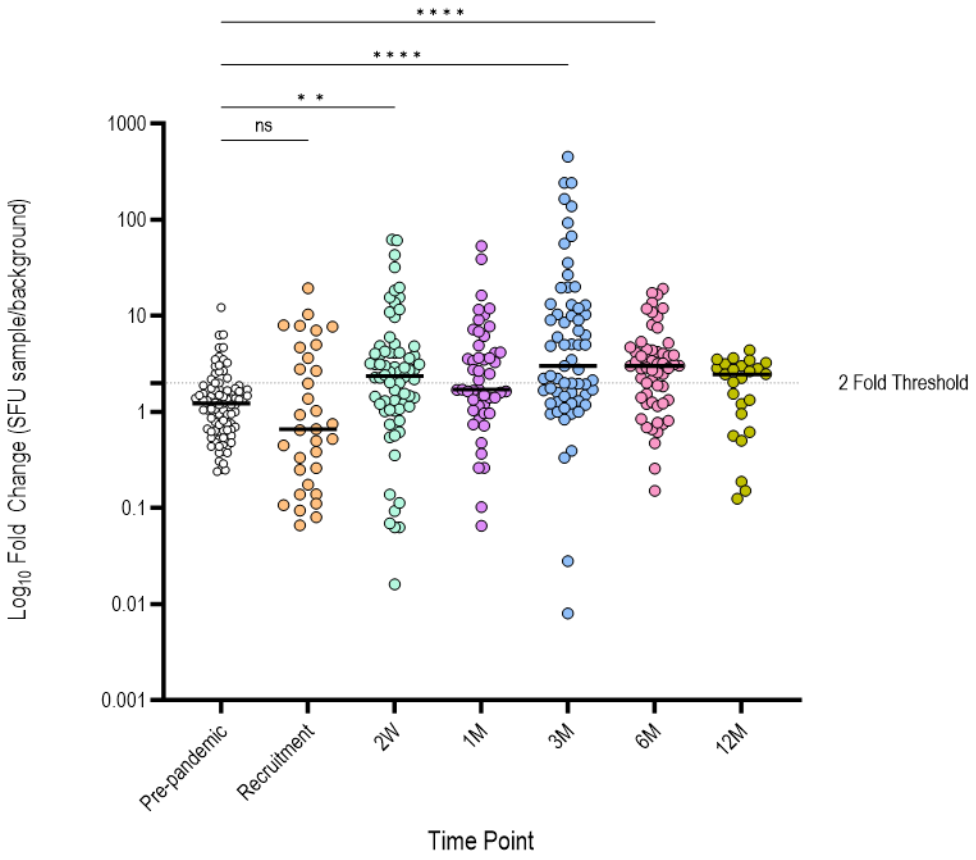


Figure 15. Reactivity to $CD8^+$ T cell Spike epitopes after vaccination and infection as measured by ELISpot IFN-γ secretion assay. Reactivity to spike is highest 2 weeks and after 3 months post-infection/vaccination. Time points are R (recruitment), 2W (2 weeks), 1M (1 month), 3M (3 months), 6M (6 months) and 12M (12 months) post-infection or vaccination. Black bars indicate the median. Statistical comparison performed by non-parametric Kruskal-Wallis using Dunn's test of Mean rank difference (*P=0.0112, **P=0.014, **** P < 0.0001).

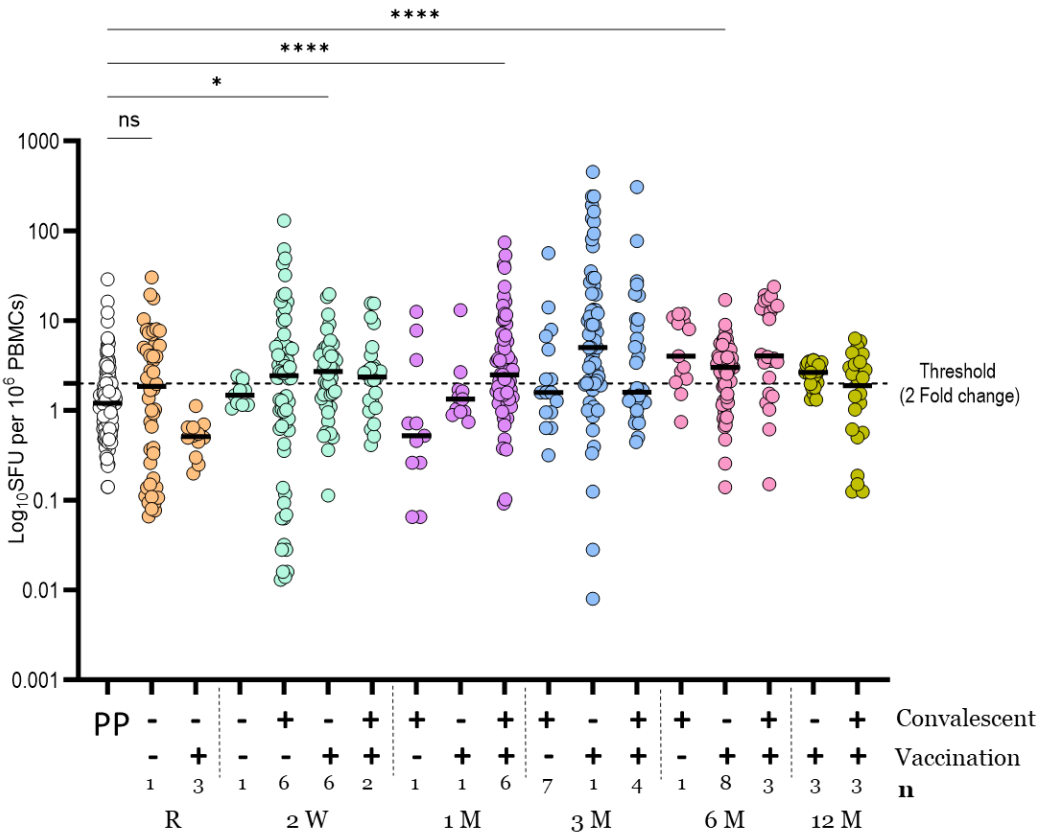


Figure 16. Reactivity to CD8⁺ T cell structural epitopes after vaccination and infection as measured by ELISpot IFN-γ secretion assay. Reactivity of PP (pre-pandemic) samples compared to 5 different time points after vaccination or infection during the Delta and early Omicron waves in 2021 and 2022. Time points are R (recruitment), 2W (2 weeks), 1M (1 month), 3M (3 months), 6M (6 months), and 12M (12 months) post-infection or vaccination. Black bars indicate the median. Statistical comparison was performed by non-parametric Kruskal-Wallis using Dunn's test of Mean rank difference (**** P < 0.0001).

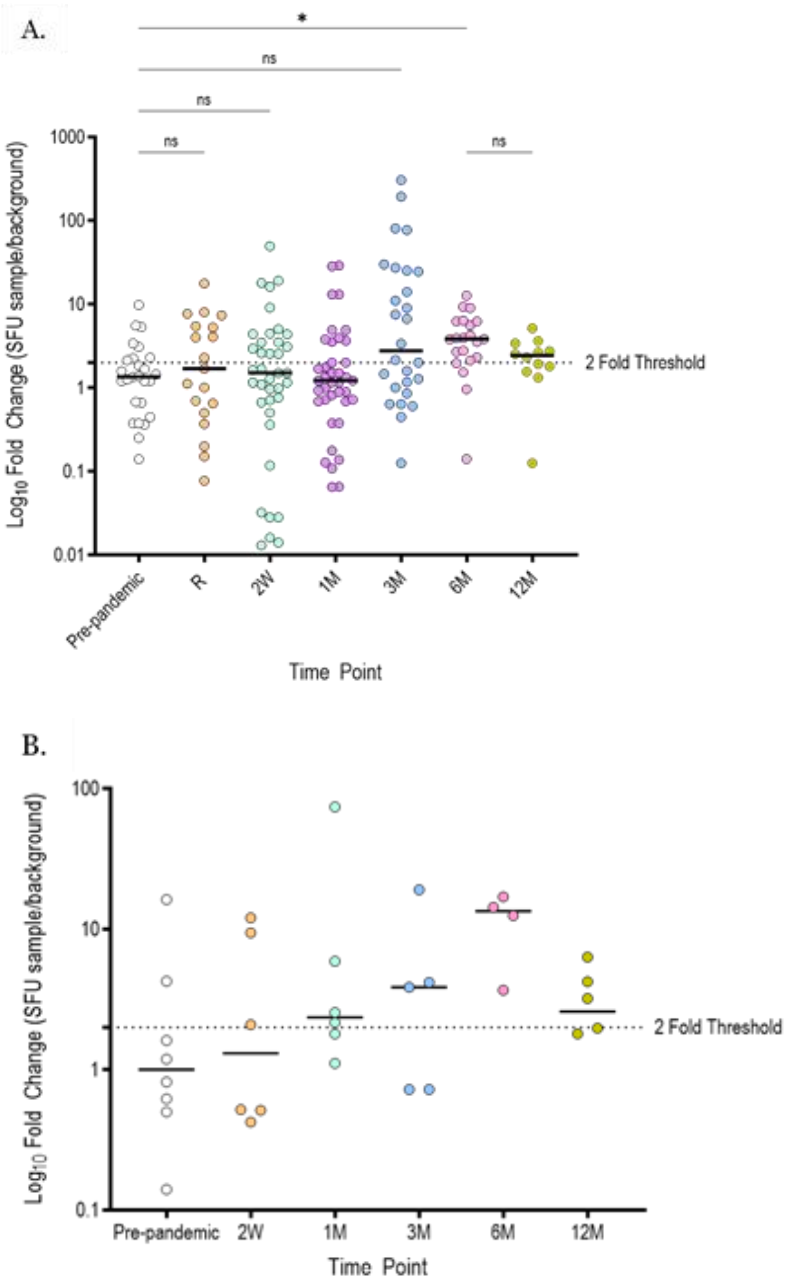


Figure 17. Reactivity to CD8⁺ T cell Membrane and Envelope peptides after vaccination and infection as measured by ELISpot IFN-γ secretion assay. A. Reactivity to membrane epitopes. B. Reactivity to envelope epitopes is non-significant between time points. Time points are R (recruitment), 2W (2 weeks), 1M (1 month), 3M (3 months), 6M (6 months) and 12M (12 months) post-infection or vaccination. Black bars indicate the median. Statistical comparisons were performed by non-parametric Kruskal-Wallis using Dunn's test of Mean rank difference (*P=0.0392), ns=not significant.

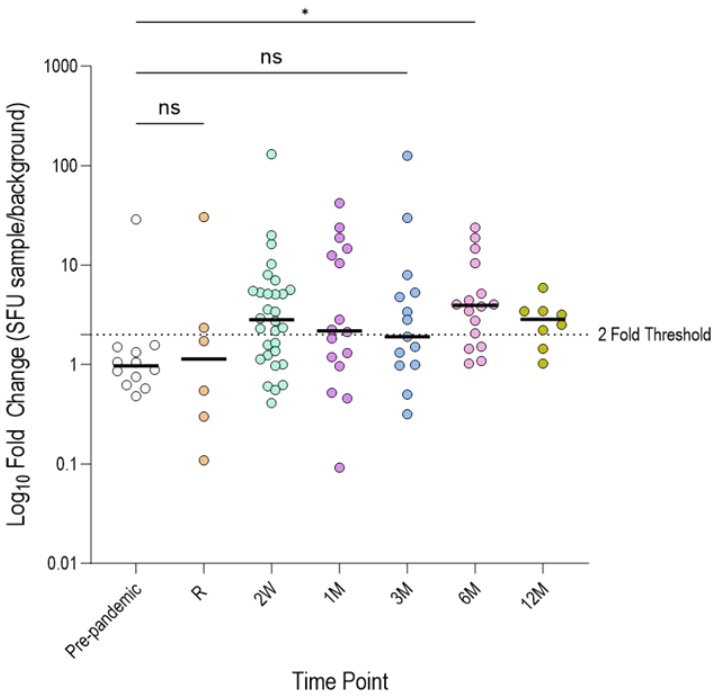


Figure 18. Reactivity to CD8⁺ T cell nucleocapsid epitopes after vaccination and infection as measured ELISpot IFN-γ secretion assay. Time points are R (recruitment), 2W (2 weeks), 1M (1 month), 3M (3 months), 6M (6 months) and 12M (12 months) post-infection or vaccination. Black bars indicate the median. Statistical comparisons were performed by non-parametric Kruskal-Wallis using Dunn's test of Mean rank difference (*P=0.0148), ns=not significant.

2.2.6 Sero-reactivity to S and N proteins

To evaluate the relation between sero-reactivity to viral proteins and T cell epitope reactivity, we determined sero-reactivity to the nucleocapsid (N) structural protein by Rapid Antigenic Protein in Situ Display ELISA (RAPID-ELISA) at each time point of sample collection. Glutathione S-Transferase)-tagged proteins are expressed using a mammalian expression system based on human HeLa cell lysate IVTT (in vitro transcription-translation) system and then captured on a 96-well plate to interrogate sero-reactivity from donor plasma. Although CD8⁺ T cells do not enhance antibody recognition of their antigens or help in affinity maturation, individuals who recovered from acute

SARS-CoV-2 infection with mild or outpatient disease had a response that included $CD4^+$ and $CD8^+$ T cells as well as antibody production [85]. We hypothesized that individuals with mild or asymptomatic infections should have a strong T-cell response accompanied by a good antibody response. We find that most donors were sero-reactive to the N protein of all HCoVs except for SARS-CoV and MERS-CoV. When compared to pre-pandemic donors, sero-reactivities were higher for SARS-CoV-2 and for SARS-CoV for the ASU cohort. Sero-reactivity to the other HCoVs was not significantly different between pre-pandemic and ASU groups, confirming that most people have been infected with most common cold CoVs (Fig. 19). Of interest, pre-pandemic donors do not have antibodies against the N protein of SARS-CoV, however, one year after the pandemic most donors are sero-reactive and the median reactivity is almost 7 times higher than for unexposed donors.

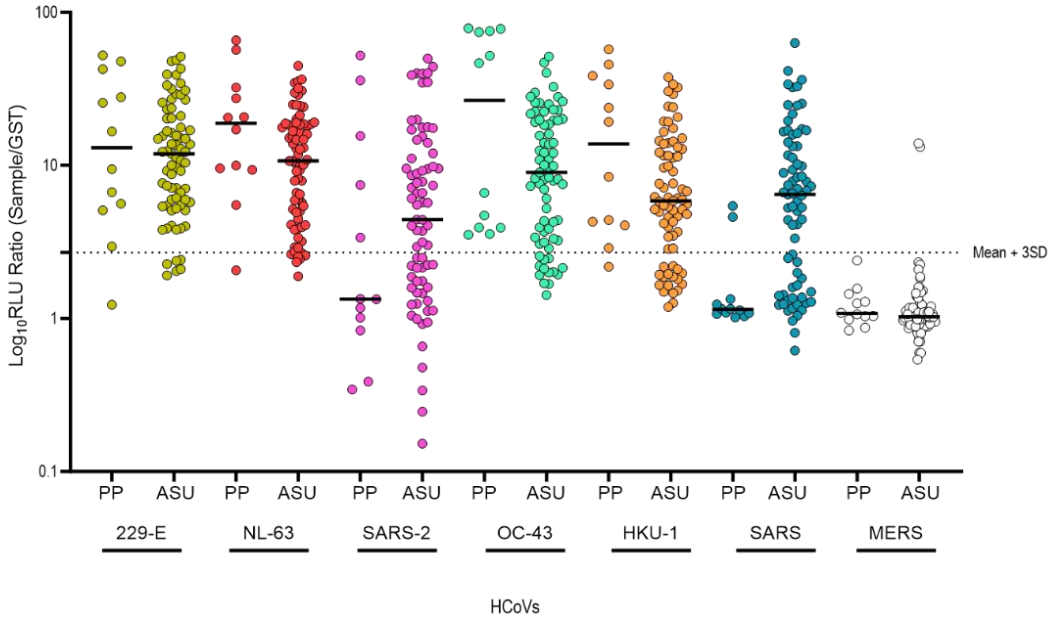


Figure 19. Sero-reactivity to the N protein in pre-pandemic and ASU cohorts. Antibodies against N protein were assayed with a RAPID-ELISA test. Samples with reactivity above the negative control plus 3 standard deviations are considered reactive. The highest MERS-CoV seroreactivity value and standard deviation were used as the threshold value for reactivity.

2.3 Discussion

2.3.1 EnsembleMHC predicted peptides are immunogenic

The current study identifies highly conserved and immunodominant structural SARS-CoV-2 $CD8^+$ T cell epitopes recognized up to 12- and 15 months post-infection and vaccination respectively. To define the targets of $CD8^+$ T cell adaptive responses in vitro, we screened a set of peptides obtained from a novel consensus algorithm, EnsembleMHC, that predicts peptides based on the molecular and biochemical features of antigen processing, presentation and binding to the HLA (human leukocyte antigen) molecule. Specifically, EnsembleMHC integrates 7 different algorithms that provide predictions based on peptide binding affinity, stability of the pMHC complex, structural characteristics of the HLA peptide-binding pocket, proteasomal degradation, and presentation on the cell surface [43]. The combination results in a reduction of false positives and provides a confidence score or $peptide^{FDR}$ for the top 0.5% of top binders. The $peptide^{FDR}$ score is a snapshot of the characteristics that define, in part, an immunogenic peptide: high binding affinity and probability of being presented on the cell surface by the MHC class I complex and finally, a high probability of being processed by the immunoproteasome which has a bias for producing antigenic peptides with C-terminal hydrophobic residues perfect for MHC class I binding. Although EnMHC has the same faults as the algorithms it is based on, being trained on the most predominant HLA class I alleles, underrepresentation of alleles from certain ethnicities, the use of pan-MHC binding prediction platforms like NetMHCpan4.0 and PickPocket ameliorates this problem. NetMHCpan4.0 uses a neural network to link known HLA class I allele binding profiles with amino acid-specific features of the peptide binding groove and extrapolate this information to determine binding affinities of uncharacterized HLA class I alleles [35, 47]. For instance, NetMHCpan-4.0 has a positive predictive value (PPV) of about 54%,

when using a data set of ~16,000 mass spectrometry-identified MHC-ligands and more than 1,250 validated T cell epitopes from the IEDB (immune epitope database). When used as part of a consensus model like EnMHC, on a set of 10 patient-generated tumor samples, the PPV under restrictive binding affinity thresholds (BA≤50nM, FDR≤5%) was 3.4 fold higher than when using the individual components [43].

In the present study, we find that 82% (76/93) of tested peptides predicted with EnMHC are immunogenic, highlighting the strength of using a consensus platform to find MHC class I ligands. The ability of EnMHC to find immunogenic peptides may be higher if we consider that most of the non-immunogenic peptides in our cohort (14/17) were tested only at one-time point and in one donor. In fact, 14/17 epitopes have been previously validated as immunogenic in other studies which brings the percentage of correctly identified epitopes to 97% (Table 10). We compared the efficacy of EnMHC to predict immunogenic epitopes with the class I immunogenicity tool by IEDB [22] and found that 65/108 peptides have a positive score predictive of reactivity, missing 10 (13%) true epitopes (Appendix D). Studies using a larger number of epitopes and in the context of different infectious diseases may provide further insight into the accuracy of this platform in finding T-cell targets. Yet, EnMHC has been validated in silico using confirmed T cell epitope data sets from the Dengue and Hepatitis B genome polyproteins and the HIV POL-GAG protein confirming the ability of EnMHC to make MHC ligand predictions over any of the 7 algorithms individually [43].

Mass spectrometry (MS) is the only immunopeptidomics method that can confirm cell surface presentation and binding between a predicted peptide and an MHC molecule [86]. Thus, the current project refers to the observed binding capacity as the number of peptides that were reactive when PBMCs were stimulated by a set of peptides predicted to be restricted to the donor's MHC genotype [87-89]. Although it is not a confirmatory result

of HLA restriction and presentation, it provides evidence that supports the predictions made by EnMHC and MHC-Flurry predictive algorithms. When comparing the predicted binding capacity with the observed stimulatory capacity of individual HLAs, there was no significant difference (Wilcoxon test, P=0.1226). However, the number of immunogenic epitopes was lower than the predicted number, probably due to the low number of available cells per donor that in turn, decreased the number of peptides that we could test. For instance, in 5 donors (ASC-7, 59, 61, 63, and 121) about half of the predicted peptides could not be tested due to an insufficient number of cells. Nevertheless, in 9/19 of the ASU donors, all the tested peptides were immunogenic while in three individuals (ASC 121 and 7), only half of the tested peptides were recognized. Overall, the predicted median number of recognized structural peptides per donor is 12, after testing reactivity, a median of 10.2 epitopes were recognized per donor. Tarke et al report that on average, an individual can recognize about 17 $CD8^+$ T cell epitopes from the SARS-CoV-2 ORFeome when considering HLA class I A and B alleles [52]. In comparison, our reported value of 10 epitopes per individual seems high when considering that this number represents only structural epitopes. However, our study used 10-day in vitro stimulations known to amplify rare T cell populations and have higher sensitivity than multiepitope ex vivo stimulations that last 24 hours as in the study by Tarke et al [52]. In addition, we may be underestimating the number of possible SARS-CoV-2 epitopes because the pool of peptides being considered does not include peptides that can be generated from non-canonical open reading frames identified in SARS-CoV-2 which are immunogenic and have been validated by LC-MS/MS as part of the SARS-CoV-2 ligandome [88, 89].

It is unclear if in the context of SARS-CoV-2, there is an effective binding capacity size at the personal or population level that is protective against severe COVID-19 disease or that may be correlated to mild or asymptomatic SARS-CoV-2 infection. To answer that

question, it would be necessary to have a higher number of donors, screen non-structural as well as structural peptides, and have a wide diversity of ethnicities and HLA class I genotypes in addition to clinical data like symptom severity. Future studies that include data mining of HLA-typed cohorts associated with disease outcomes will clarify the importance of binding capacity as a criterion for evaluating the disease susceptibility of individuals or populations in addition to individual HLA class I and II allele associative studies.

As a predictive algorithm, EnMHC combines the best features of other consensus platforms like NetMHCpan and MHC Flurry, it is trained on BA and EL data and has a better PPV when used under restrictive parameters. Although EnMHC was trained and benchmarked with decoy peptides, our in vitro validation did not use negative controls such as peptides restricted to different HLAs due to the low cell number available for testing.

Table 10. List of peptides that were non-immunogenic in this study but were validated in other SARS-CoV-2 $CD8^+$ T cell epitopes studies.

Peptide	Protein	Peptide Sequence	Validation Method	HLA-I Restriction	Reference
65	E	YVYSRVKNL	Multimer staining	C06:02	Saini, 2021; Snyder, 2020
77	E	LVKPSFYVY	Multimer staining	C07:02	Saini, 2021
13	M	ATSRTLSYYK	AIM, multimer staining	A30:01	Tarke, 2021; Titov, 2022
80	M	RLFARTRSMW	AIM	A32:01	Tarke, 2021
49	N	MEVTPSGTW	Multimer staining, AIM	B44:03	Titov, 2021, Schulien, 2021
12	S	ASANLAATK	ICS	A11:01	Jin, 2021
20	S	QYIKWPWYIWLGF	A longer, by AIM	A24:02	Shimizu, 2021
37	S	IPFAMQMAY	AIM	B35:01, B51:01	Tarke, 2021
38	S	IPIGAGICASY	HTMA	Unknown	Snyder, 2020
39	S	LPFFSNVTW	LPFFSNVTWFHA	not reported before	
50	S	QELGKYEQY	AIM	B44:02, B44:03	Tarke, 2021
52	S	IPTNFTISV	AIM	B07:02, B51:01	Tarke, 2021
53	S	LPLVSSQCV	AIM	B51:01	Tarke, 2021
67	S	KIADYNYKL	Multimer staining	A02:01	Saini, 2021
68	S	VLNDILSRL	AIM	A02:01	Tarke, 2021
83	S	DISGINASV	not reported		not tested
98	S	HADQLTPTW	Multimer staining	A02:01, B07:02	Weingarten-Gabbay, 2021
101	N	LPNNTASW	AIM, HTMA	B07:02	Tarke, 2021; Snyder, 2021

2.3.2 Immunodominant epitopes are highly conserved among VOC

Using EnMHC, we obtained a list of 658 peptides from the entire ORFeome of SARS-CoV-2 and prioritized structural peptides for testing for two main reasons: 1) variability in the number of structural ligands per genotype (per donor) was smaller and the repertoire size more manageable (n=108), and 2) previous reports of SARS-CoV immunogenicity indicated that non-structural proteins were not as reactive or immunodominant as structural proteins [75]. However, in the last three years, a growing body of evidence shows that a few nsps (e.g. nsp12 and nsp3) are as immunodominant as structural proteins [90]. Still, we find that the breadth of response for the ASU cohort includes S as well as M, N, and E epitopes. More comprehensive studies using mega peptide pools of the entire SARS-CoV-2 ORFeome, mapped 26% of reactivity to spike, 12% for N, and 22% for M [74]. Since we only tested structural proteins, our percentages are higher for the ASU cohort, 56% of the response was directed to S, 26.7% to M, 13.3% to N, and 4% to E. Spike immunodominance may be explained by the correlation between the number of positive responses per protein and the size of the source protein and it highlights the validity of using the spike as the main immunogen in most of the current vaccination models. However, the spike is the structural protein most susceptible to mutations, and constant viral evolution makes it essential to include highly conserved targets from other structural and non-structural proteins in the next generation of vaccines to increase resilience to new VOC.

We have found a set of 15 epitopes that are highly immunogenic and conserved among VOC including epitopes from the S, E, M, and N proteins all within regions of low mutation rates [91, 92]. After mapping immunodominant S epitopes, only two epitopes have mutations associated with the Omicron variant, both are located within the RBD of the spike and the rest are located at the N-terminal domain and S2 region of the spike, both

regions with low non-synonymous mutation rates [92]. The population coverage tool predicts a minimum of 1.55 epitopes per HLA combination recognized by 90% of the population and an average of 3.4 epitopes per HLA combination recognized by the global population, making this group of immunodominant epitopes high-value targets to include in next-gen vaccine models [78]. By including epitopes from other structural proteins, next-gen immunization platforms have the potential to closely mimic exposure to antigens during natural infection.

From the total 108 structural peptides predicted with EnMHC, 17 peptides had at least 1 amino acid change in one or more VOC. Although we could not test all variant epitopes due mostly to HLA restrictions, from the 10 variant peptides we tested, 9 were immunogenic in the same donor. Of those, 3 were mutated at P4-6 positions and were still predicted to bind to their predicted HLA. An immunodominant spike epitope GVYFASTEK^{WT} had a change at anchor residue P7 from a hydrophilic threonine to an aliphatic/hydrophobic isoleucine residue (GVYFAS**I**EK^{O}) increasing the magnitude of response 4-fold. This peptide is restricted to the A3 supertype that prefers Arg or Lys residues at P7 which could explain the increase in reactivity from Thr to Ile since Ile is an isomer of Leu [93]. Overall, alignment of the immunodominant epitopes with VOC shows high biochemical conservation. An epitope from the RBD of S, NSASFSTFK^{WT} had several changes found in two different Omicron subvariants (Table 8), we tested both variants (only once) and found that compared to the ancestral version (WT), reactivity was increased by 5.8- and 9.5-fold for $_{370}$N**LA****P**F**F**TFK$_{378}$ and $_{370}$N**F**A**P**F**FA**FK$_{378}$ respectively. This peptide is restricted to the HLA-A*68:01 that belongs to the A3 HLA superfamily that accepts small and aliphatic residues in the B pocket and aromatic and basic amino acids in the F pocket. Both variants have a change at the P2 residue from S which is a hydrophilic residue (not preferred at P2) to L or F. Leucine is aliphatic but Phenylalanine is quite

bulky, perhaps displacing or the bulge produced by the P4-P6 residues and making the peptide more reactive. The increase in reactivity could be due to the change from an S at P4 and P6 into a P and an F residue, making the area of contact with the TCR bulkier. The only way to determine if any of these suppositions are correct is to crystalize the pMHC complex and a cognate TCR to characterize the actual occupation of the peptide binding groove and the interaction with the TCR.

Due to the small size of our sample set, we are not able to discern if amino acid changes present in VOC lead to viral evasion of T cell-mediated immunity. A few studies have found that effector and memory T cell compartments from both CD4 and CD8 T cells are maintained in convalescent and/or vaccinated individuals but there is a decrease of about 50% in either effector or memory T cell reactivity when compared to ancestral or Delta variants [94]. However, the same study found that individuals with low or no Omicron antibody neutralization titers had measurable effector and memory CD8 and CD4 T cell responses to Omicron spike peptides. An independent study comparing 10 different VOCs, found a decrease in cytokine production upon activation with Omicron peptides when using intracellular cytokine staining (ICS) and a decrease in the number of responders using an AIM assay [95]. Loss of T cell recognition may be caused by lower binding affinity to the cognate HLA, decreased recognition of the pMHC by the TCR, or changes in antigen processing and presentation due to the virus itself.

2.3.3 Vaccination increases reactivity to spike and other structural SARS-CoV-2 proteins

Most studies that look at epitope-specific T-cell responses after vaccination focus on the spike protein. In addition to the spike, we tested other structural proteins to determine spike-centered vaccines like mRNA-1273, BNT162b2, and Ad26.COV2.S changed immunogenicity to other structural proteins and characterized differences between

convalescent unvaccinated donors and among different vaccine platforms. Although our sample set was small, we were able to make a few observations. First, the breadth and magnitude of response are increased after vaccination as in the case of UU donors at the time of recruitment. After vaccination, reactivity increases by about 25 to 56-fold 2-3 months post-vaccination (Fig. 16, Table 11) and one donor (AS-25) recognized 10/11 tested peptides 3 months post-vaccination. In agreement with the previous observations, a study on vaccine efficacy shows a 10-fold increase of spike-reactive IFN-γ secreting cells 2 weeks after vaccination

Table 11. Effect of vaccination on reactivity to CD8+ T cell peptides as measured by ELISpot IFN-γ secretion assay.

	UU to UV				CU to CV		CU	
Time Point	R	2mpv	R	3mpv	R	1mpv	R	3mpi
Donor	ASC-12	ASC-164	ASC-25	ASC-177	ASC-9	ASC-118	ASC-8	ASC-117
Median Reactivity	3.201	80.273	0.124	7.000	0.630	14.700	2.400	1.580
Fold change	25.077		56.452		23.333		0.658	
Fraction Reactive Peptides	0.650	1.000	0.000	0.909	0.214	0.750	0.625	0.353
# Reactive Peptides	13	15	0	10	3	9	10	6
Total Tested	20	15	16	11	14	12	16	17
Proteins recognized	S, M, N	S, M, N	S, M, N, E	S, M, N, E	S, M, N, E	S, M, N, E	S, N, M	S, N, M

R: Recruitment; mpv: months post-vaccination; mpi: months post-infection. UU: Uninfected Unvaccinated; UV: Uninfected Vaccinated; CU: Convalescent Unvaccinated; CV: Convalescent Vaccinated. S: Spike; M: Membrane; E: Envelope; N: Nucleocapsid.

with an attenuated SARS-CoV-2 virus [96, 97]. Second, after vaccination, the reactivity is increased for all structural proteins not only for spike (Fig. 16-18). We find that immunogenicity to M, N, and E epitopes is enhanced after vaccination with all three vaccine models. For example, reactivity to the immunodominant E epitope FLAFVVFLL is highest in a convalescent donor vaccinated with the BNT162b2 vaccine three months before sample collection. The previous suggests that vaccination with spike-centered formulas enhances immunogenicity to other structural SARS-CoV-2 proteins. Responses are also maintained up to a year after vaccination and/or infection. For instance,

NRFLYIIKL, an epitope from the membrane protein, was highly reactive in up to 6 different vaccinated donors (n=5/6). The strongest responses to this epitope came from two donors, ASC-118 who received the second dose of mRNA-1273 vaccine 29 days before sample collection, and another convalescent donor who had received an Ad26.COV2.S (J&J) vaccine 12 months prior (ASC-461). This pattern of enhancement is known as antigenic spread, determinant spread, or antigen cascade. In the context of cancer, cell death of tumor cells releases tumor-associated antigens (TAAs). Antigen-presenting cells (APC) uptake TAAs and, induce a B or T cell response that includes the production of cytokines and chemokines that further activate surrounding cells to recognize other antigenic targets in neighboring cells. It is likely that vaccination or the combination of prior infection and vaccination leads to a response against antigens (epitopes from N and E) that are not part of the vaccine, but only in previously infected individuals.

Reactivity to spike epitopes increased after 2 weeks post-infection or vaccination and remained elevated above the 2-fold threshold over background after 3 months and was sustained up to 15 months post-vaccination and 8 months post-infection (unvaccinated donor). Reactivity was enhanced to all structural proteins except for the M protein which had a lower median than the other structural proteins except at three months. It is unclear why membrane has a lower median of reactivity. Some peptides have reactivities that are over 40-fold over the background while others barely react 1-fold over the background during the same time point for the same donor. The current study looks at the impact of VOC on immunodominant epitopes, it is not meant to be an exhaustive look at the ORFeome of SARS-CoV-2 that is affected by amino acid changes present in VOC. However, we find that the immunodominant epitopes we have characterized in this study are not likely to be affected in future VOC and show sustained reactivity after >12 months post-vaccination. We also observed an increase in background reactivity 6 months after

vaccination and after booster application. A similar phenomenon has been described in individuals who had an elevated level of $CD4^+$T cell activation due to recent illness and background for ELISpot IFN-γ [98]. It is also possible that the increase in reactivity was due to higher chances of exposure since the 6- and 12-month samples were collected during the Fall of 2021 spike in Delta infections and the spread of the new VOC Omicron during the Spring of 2022.

2.4 Limitations and Future Directions

Although the current study validates the immunogenicity of peptides restricted to 31 HLA class I alleles that are predicted to cover >99.56% of the population, our cohort was not representative of African American (AFA) or Native American (NA) populations or minorities in general. Since minority populations have fared the worst from this pandemic, it is of great importance to study T cell ligand repertoires for these populations and determine how the current vaccine models may be improved or benefit them. On the other hand, one of the strengths of this study was the use of samples from young individuals who either did not have any reported SARS-CoV-2 infections or who were convalescent from mild or asymptomatic COVID-19. Although all recovered donors had outpatient disease that lasted 15-30 days, we did not have follow-up information on re-infections or exposures that could have explained some of the spikes in reactivity that were observed at 3- and 6-months post-infection and/or vaccination. Another drawback of this study was the lack of information on the history of common cold or influenza infections that could have been useful to further understand pre-existing immunity or cross-reactivity. Symptomatology related to SARS-CoV-2 infection was not gathered during sample collection impeding any efforts to understand trends in asymptomatic vs symptomatic COVID-19 disease. The present cohort was mostly male and healthy which gives us a peek at what healthy responses look like, however, we don't have access to

samples from donors with more severe disease or older populations to make a comparison. Finally, data collection may be more productive and save reagents if it is done according to a timeline set either by the data of infection or vaccination. Donors that enrolled had to donate samples at the same time intervals regardless of how much time had passed after infection or vaccination which confused the results and required a complicated tabulation that was lengthy and laborious to place each sample in approximately the same time frame after infection/vaccination.

2.5 Materials and Methods

2.5.1 Isolation of peripheral blood mononuclear cells (PBMCs) and plasma

2.5.1.1 Healthy Unexposed Pre-pandemic donors

Peripheral blood mononuclear cells (PBMCs) and plasma samples were isolated from healthy donors during June 2017 and June 2019 before the start of the SARS-CoV-2 pandemic. Samples were obtained from Blood Centers of the Pacific from consented donors. 12 samples were processed according to the institutional review board (IRB) of Arizona State University (ASU), IRB #00006056. The cohort had a sex ratio of and an age between years). SARS-CoV-2 seronegativity was determined by Access ELISA as described below. Demographic information and HLA phenotype are listed in Table 4.

2.5.1.2 Healthy and COVID-19 convalescent donors

Blood from healthy seronegative and COVID-19 convalescent donors was collected and processed according to approved protocols from the ASU student study IRB (#00011804). Study participants belong to the ASU community (n=19) and may or may not be Arizona residents, the range of cohort age is between 19 to 51 years old (median: 25 y/o), and a sex ratio of females to males of 4 to 15 (Table 4). Seroprevalence at the time of enrollment was determined by Access Immunoassay against the SARS-CoV-2 spike (described below).

2.5.1.3 Collection of Peripheral blood mononuclear cells (PBMCs), plasma, and HLA typing

PBMCs and plasma from whole blood were collected using CPT tubes (BD Vacutainer CPT Cat#362753) following the manufacturer's instructions. Briefly, PBMCs are separated from plasma and platelets during 5 consecutive centrifugations done at 15°C using different speeds and times. PBMCs are cryopreserved in heat-inactivated FBS (fetal bovine serum) containing 10% DMSO (GIBCO) and stored in liquid nitrogen until needed. Sera was stored at -80°C in 1.5 mL Eppendorf tubes (3 aliquots of 1 mL each). Samples were HLA typed using exome sequencing after isolation of RNA from 1 million PBMCs pelleted at 330 RCF for 8 minutes. RNA was isolated using the RNAqueous Kit (AM1912), concentrated and buffer exchanged using Zymo RNA clean concentrator (R1018). The final sample was submitted to the Genomics Core at Arizona State university (ASU) for TapeStation analysis (Agilent D5000 high sensitivity 5067-5592) and peptide library construction using KAPA's mRNA HyperPrep kit (KAPPA KK8580). The mRNA was captured with Magnetic oligo-dT beads and sheared down to 250-300 base pairs using heat and magnesium. Illumina-compatible adapters (IDT #00989130v2) were ligated and excess was removed using KAPA pure beads (KAPA KK8002) then amplified with Kapa's HIFI enzyme (KAPA KK2502). Fragment size and quality were verified with Agilent TapeStation and then quantified by qPCR (KAPA KK48835) on a Thermo Fisher Scientific QuantStudio 5 before multiplexing and sequencing on an Illumina NovaSeq6000 2x150 flow cell at the University of Colorado, Anschutz Medical Campus.

2.5.2 PCR Saliva Test

Detection of SARS-CoV-2 RNA was performed at the CLIA-certified ASU Biodesign Clinical Testing Laboratory. RNA was isolated and purified from saliva samples collected

from donors at the time of blood collection. About 300 uL of the donor's sample was mixed with proteinase K and RNA was extracted using the Qiagen RNeasy mini kit (Qiagen Cat# 74106). Samples were spiked with the MS2 phage RNA as an internal control (proprietary information from Thermo Fisher Scientific-Cat# A47814). Multiplex RT-qPCR was performed following the manufacturer's instructions. The TaqPath COVID-19 combo kit was used to detect ORF1ab, N, and S genes using MS2 as the internal control and a TaqPath positive control kit that contains 10^4 copies of SARS-CoV-2/uL. Ct values for SARS-CoV-2 N and S genes were determined using FastFinder software (V.3.300.5).

2.5.3 Serology

2.5.3.1 SARS-CoV-2 RBD-ELISA of Pre-pandemic and ASU cohort samples

HCoV RBD proteins: SARS-CoV-2 RBD from 7 different coronaviruses (HKU-1, OC-43, SARS-CoV-2, SARS-CoV- MERS, 293E, and NL63) were produced as secreted proteins using the Expi293™ expression system kit (ThermoFisher A14635). Briefly, the RBD of each HCoV was cloned into the pcDNA3.4 expression vector flanked by a signal peptide (MDAMKRGLCCVLLLCGAVFVSP) of the human tissue plasminogen activator at the N-terminus and a C-terminus $(GGGGS)_3$-Halo-3xFLAG tag. Endonuclease-free plasmid DNA was transfected into Expi293F cells using the ExpiFectamine293 Transfection Kit (Gibco). Cells were cultured in Opti-MEM-I reduced Serum Medium (Gibco) in an orbital shaker at 37°C with 8% CO2. Four and a half days post transection, the cells were pelleted, and the supernatant was clarified by centrifugation for 20 minutes at 4,000g and 4°C. The supernatant was spin-concentrated and filtered using 0.22 um Stericup, aliquoted, and stored at -80°C. Protein expression was verified by in-gel fluorescence assay. RBD supernatants were mixed with HaloTag Alexa Fluor 660 Ligand (Promega G8471) and allowed to bind for 30 minutes at room temperature. The samples were mixed with XT gel loading dye (Bio-Rad) and boiled for 10 minutes at 95°C. The samples were loaded on a 4-

20% SDS-PAGE gel and a fluorescent signal was detected with a laser scanner platform (Typhoon trio).

2.5.3.2 SARS-CoV-2 specific IgM/IgG Quantitative Detection by Access Immunoassay of ASU Study Samples

Detection and quantification of IgM and IgG antibodies against SARS-CoV2 RBD (receptor binding domain) was performed by the CLIA-certified lab from the Biodesign Institute at ASU. The serum was separated from whole blood samples collected from the ASU Student Study participants as previously described. Serum samples were aliquoted and only thawed once to perform the access Immunoassay using the Beckman Coulter protocol and analyzer (Cat# C58961 and #C58957). Briefly, the *Access* assay uses paramagnetic particles and a chemiluminescent assay to detect and quantify IgG and IgM antibodies against the SARS-CoV-2 RBD. The paramagnetic particles were coated with a mouse anti-human IgM or IgG antibody that was incubated with the prediluted donor sample. The unbound material was washed away, and the particles incubated with recombinant SARS-CoV-2 RBD protein tagged with an alkaline phosphatase tag. After a second wash, a chemiluminescent substrate is added for the quantification of light produced by the machine's luminometer. A cut-off value is determined during calibration of the instrument (Beckman Coulter UniCel Dxl 800) and results are interpreted as negative (<10 AU/mL) or positive (≥10 AU/mL).

2.5.3.3 Detection of Human CoV N-specific IgG by Rapid-ELISA

The presence of antibodies against the nucleocapsid protein of Human Coronaviruses (NL-63, 229E, OC43, HKU-1, SARS-CoV, SARS-CoV-2, MERS-CoV) was detected by chemiluminescence. Briefly, a 96-well plate is coated with a Goat polyclonal anti-GST (glutathione S-transferase) at 4°C overnight, washed (0.2% PBST), and blocked for 1.5 hours at room temperature with a 5% milk PBST solution. After 5 washes the plate is

coated with recombinant CoV nucleocapsid tagged with GST (see below) for 1 hour at RT (room temperature). After washing, pre-diluted donor sera (1:100 serum to block serum buffer-*E. Coli* lysate) is added for an hour incubation at RT. Plates are then probed with a goat anti-human IgG antibody conjugated to HRP (horseradish peroxidase) (Jackson ImmunoResearch #5150035-062) for 1 hour at RT. Using a Promega GloMax luminometer, peroxidase activity is detected at 425 nm wavelength, 1-5 min after adding the substrate solution (Thermo Sci. #37074).

Cell-free expression of recombinant human CoV nucleocapsids: Nucleocapsids (NC) tagged with GST from different human coronaviruses (NL-63, 229E, OC43, HKU-1, SARS-CoV, SARS-CoV-2, MERS-CoV) are expressed using a cell-free system or IVTT (In vitro transcription-translation). A HeLa lysate master mix (Thermo Scientific) was combined with the antigen plasmid (e.g. NL-63 NC pANT7_cGST) and the reaction was left in a 30°C incubator for 1.5 hours. The resulting antigen is diluted 1:100 in a 5% milk PBST solution. All plasmids were purchased from DNASU (Center for personalized diagnostics, ASU) [99].

2.5.4 Epitope prediction and peptide selection

We used EnsembleMHC (https://github.com/eawilson-CompBio/EnsembleMHC-Covid), a novel computational tool that integrates peptide binding affinity, and peptide-pMHC complex stability with peptide-binding pocket structural features and the probability of antigen presentation [43]. Using the ancestral variant (MZ_468053) we obtained 658 CD8-T cell epitopes from the SARS-CoV-2 proteome restricted to the most common HLA alleles in the world and prioritized testing the structural peptides (n=108) restricted to the HLA types in our student study cohort.

2.5.5 Amino-acid Conservation in homologous CoVs and SARS-CoV-2 VOCs

Immunogenic epitopes (after ELISPot and AIM analysis) were matched to homologous non-pathogenic coronavirus NL63 (NC_005831.2), 229E (KY_684760.1), OC43 (AY_391777.1), HKU-1 (AY_884001.1), SARS-CoV (AY_274119.3), and MERS(JX_869059.2) and emerging variants B.1.1.7 (MZ_202178), B.1.137 (MZ_468007), P.1 (MZ_202306) and B.1.617.2 (MZ_468047) and B.1.1.529 (BA.1-BA.5; OL_965559). Alignments were done using the Clustal Omega tool from EMBL-EBI (www.ebi.ac.uk/services) [100] and visualized using JalView [101].

2.5.6 SARS-CoV-2 ancestral and variant peptide synthesis

We used 108 MHC Class-I -restricted epitopes (ProImmune, UK. >80% purity) and tested individual peptides according to the HLA type of the donor and the number of live cells after thawing. We used a CEF pool (CMV, EBV, and Flu viral epitopes) (ProImmune, UK) as a positive control and two negative controls, untreated cells and PBS with 5% DMSO.

2.5.7 ELISpot assay

Enzyme-linked immunosorbent assay (ELISpot) was performed as previously described [102]. Briefly, a 96-well multi-screen plate (Millipore) was seeded with 5ug/well of anti-IFNγ (Mabtech) capture antibody diluted in PBS overnight at 4°C. The peptide-pulsed PBMCs were transferred to the ELISpot plate and incubated at 37°C for 48 hours. The plate was washed with ELISpot buffer (PBS + 0.5% FBS) and 1µg/mL of anti-IFNγ (Mabtech) secondary detection antibody was added and left at room temperature for 2 hours. Streptavidin ALP conjugate (1µg/mL, Mabtech) is added and the plate is incubated for 1 hour, and washed with ELISpot buffer. The plate was developed with detection buffer (33µL NBT, 16.5µL BCIP (Promega), in 100 mM Tris-HCl pH 9, 1 mM MgCl2, 150 mM NaCl) for 8-10 minutes and the reaction was stopped by rinsing with water 5 times and allowing the plate to dry for up to two days. Spots were read using the AID ELISpot reader

(Autoimmune Diagnostika GmbH). The average number of spot-forming units was calculated for each tested pool or individual peptide and subtracted from the background (either untreated or PBS-DMSO control). A reactive sample was above a 2-fold threshold of SFU (spot forming units) per 10^6 PBMCs of sample over background (untreated or DMSO/PBS, whichever was lower). Each peptide and control were tested in triplicates. Samples that did not have more than 5 SFU/10^6PBMCs for the positive control were not considered for analysis. Due to the small sample size and limited number of cells, immunogenic peptides were prioritized for testing when not enough cells were available at subsequent time points.

2.5.8 Flow cytometry staining for activated-T cells (AIM).

Stimulated PBMCs will be stained for activation markers, specifically CD137/CD69 for CD8+ T cells and CD134/CD137/CD69 for CD4+ T cell markers to quantify epitope-specific T cell activation [103]. We used the Attune™ NxT Flow cytometer in conjunction with the FlowJo™ software for data acquisition and analysis.

2.5.9 Quantification and statistical analysis

All tests are stated in the figure legends. P values are specified by each table or figure.

CHAPTER 3

CD8+ T Cell Pre-Existing Immunity to SARS-CoV-2

3.1 Introduction

Respiratory viral infections like influenza virus, adenoviruses, respiratory syncytial virus (RSV), and coronaviruses manifest with an array of symptoms that can range from asymptomatic, to mild to severe and in some cases lead to death [104]. For most infections, viral titers of exposure, genetic background of the infected individual (e.g. HLA genotype), or viral-dependent co-morbidities (e.g. heart disease, diabetes) will determine the course and severity of the infection [105]. However, pre-existing immunity due to cross-reactive responses may affect susceptibility to infection and disease severity.

Documented cases of pre-existing immunity due to prior infection or exposure to related viruses can reduce susceptibility or enhance disease. For instance, secondary infections with flaviviruses like Dengue or Zika virus can be cross-protective [106, 107], or detrimental as in the case of disease enhancement by influenza T cell cross-reactivity to Hepatitis C Virus CD8+ T cell epitopes (HCV) [108]. SARS-CoV-2 shares 65% to 69% of sequence homology to common cold coronaviruses, and has a higher similarity to the pathogenic SARS-CoV (86%) [109]. This degree of similarity suggests that prior infections with CCCoVs generate an adaptive immune response that may cross-react with similar viruses like SARS-CoV-2. In fact, antibody and T cell responses are part of the response against CCCoVs but antibody levels decrease faster that T cells. A study looking into adaptive responses to SARS-CoV found that antibody levels decrease below detectable levels two to three years post-infection, while virus-specific T cells can be found up to 17 years later [110, 111].

Here, we explore the degree of cross-reactivity to SARS-CoV-2 in unexposed individuals by testing the reactivity of CD8+ T cells to viral structural epitopes and whether

that response is due to cross-reactivity to CCCoV like OC-43 and HKU-1. Since more than 90% of the adult population have antibodies against HCoVs that cause the common cold [112], there are likely T_{mem} cells that cross-recognize SARS-CoV-2 $CD8^+$ T cell epitopes. Understanding cross-reactivity in the context of SARS-CoV-2 may elucidate some of the rules that determine the vast disparity in clinical outcomes of COVID-19 disease. In addition, identifying discrete regions of structural proteins that are immunogenic in unexposed individuals, may provide insight into the cross-reactive nature of the spike protein which has implications for vaccine design and the addition of other structural targets to generate protective memory T cells against SARS-CoV-2.

3.2 Results

3.2.1 Structural Epitope repertoire in SARS-CoV-2 unexposed individuals

To determine the level of pre-existing immunity to SARS-CoV-2 and the $CD8^+$ T cell epitope repertoire in unexposed individuals, we tested recognition and magnitude of reactivity by ELISpot IFN-γ secretion assays using peptides derived from the ancestral sequence of SARS-CoV-2 (Table 2). To test the immunogenicity of structural peptides, we used peripheral blood mononuclear cells (PBMCs) collected from 12 donors between June 2017 and June 2019 before the declaration of the SARS-CoV-2 pandemic in March 2020 and well before the initial global spread during the first months of 2020. Our pre-pandemic (PP) cohort (n=12) is composed mostly of white females with a median age of 59 y/o (Table 4). Unexpectedly, 42% of donors were sero-reactive to the N protein of SARS-CoV-2 as measured by RAPID-ELISA (Fig. 19). PBMCs were stimulated for 10 days with individual peptides derived from all four structural proteins and restricted to 21 unique HLA class I alleles (Fig. 6). Long term in vitro stimulations have been used extensively to amplify low-frequency T cell responses to viral antigens and have also been used in the detection of bacterial and allergenic antigens recognized by $CD4^+$ and $CD8^+$ T

cells [28, 113]. The $CD8^+$ T cell peptides to be tested were selected from a group of 108 structural peptides predicted with the consensus algorithm EnMHC and restricted to 52 of the most common HLA class I alleles [43]. Restricted by the HLA genotypes of our cohort we tested 59/108 peptides, 3 peptides from the envelope (E), 14 from the membrane (M), 7 from the nucleocapsid (N), and 35 from the spike (S) proteins (Fig. 20).

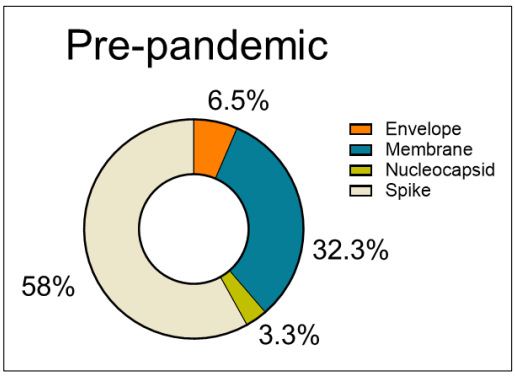


Figure 20. The proportion of structural proteins recognized in unexposed donors. A total of 59 peptides were tested, the graph shows the distribution of immunogenic peptides only. 22 peptides were immunogenic, S=12; M=8; N=1; E=1.

A total of 31 epitopes were immunogenic, that is the reactivity measured as the number of spot-forming units (SFU) normalized to 1 million PBMCs was over 2-fold that of the background response from the negative control. Each donor recognized an average of 2.5 epitopes and a median of 2, with a minimum of cero to a maximum of 6 epitopes recognized. About 63% of epitopes (14/22) were recognized by 4 donors and of those, most were from the S protein (12/22) and the M proteins (8/22). Epitopes with the highest magnitude of response originate from all 4 structural proteins, however, most were directed to the spike protein (Table 12). Four epitopes, one from each of the structural proteins had the highest magnitudes of response.

- *Spike immunodominant and immunoprevalent $_{264}$YLQPRTFLL$_{272}$*

The $_{264}$YLQPRTFLL$_{272}$ epitope from the S protein had a 12-fold over-background reactivity and was recognized by 4 different donors. This peptide is highly immunogenic in unexposed donors as well as in convalescent and vaccinated individuals (Chapter 2). It has been validated in numerous studies, a search on the Immune Epitope Database (IEDB) reveals 48 different studies have validated this epitope by immunization of transgenic mice, IFN-γ ELISpot, and ELISA [114-116]. Supported by other studies, we find that this epitope is restricted to HLA-A*02:01 and B*08:01, and additionally report that it is immunogenic in donors restricted HLA-B*11:01, HLA-C*05:01, and C*07:01. Unfortunately, this epitope is found neighboring the RBD of spike and it is mutated in recent subvariants of the Omicron VOC. Nevertheless, the immunodominance of this epitope reported in several studies may still generate cross-reactive responses because although it is mutated, we find that it is a conserved identity.

- *Nucleocapsid immunodominant and immunoprevalent $_{263}$KAYNVTQAF$_{271}$*

The $_{263}$KAYNVTQAF$_{271}$ epitope from the N protein had a 28-fold over-background reactivity and had the highest magnitude of response among structural peptides. It has been reported in four different studies that validate this epitope as part of an immunodominant response to SARS-CoV-2 infection as it was found to be immunogenic in convalescent samples by IFN-γ ELISpot/ELISA and tested in a mouse model as part of a peptide vaccine that spans globally common HLA genotypes [114, 117]. The $_{263}$KAYNVTQAF$_{271}$ epitope has been validated to bind to HLA-B*35:01 and B*51:01. Here, we show that it can be presented in individuals restricted to HLA-B*57:01, C*12:02 and C*06:02.

- *Envelope immunodominant $_{20}$FLAFVVFLL$_{28}$*

This peptide has been validated as immunogenic in vivo after immunization of a transgenic mouse model and confirmed to be restricted to HLA-A*02:01 [118]. It stimulates the proliferation of $CD8^+$ T cells in vitro as quantified using a carboxyfluorescein succinimidyl ester (CFSE) and ICS assays [119] and by tetramer binding [116]. The last two studies were done in convalescent individuals and in Chapter 2, we characterize responses against this epitope in both vaccinated and convalescent donors, making $_{20}$FLAFVVFLL$_{28}$ a high-value non-spike target for vaccine or TCR-mediated therapies.

Table 12. Reactive epitopes in pre-pandemic samples

Pep Seq	PROTEIN	# Donors	Highest Magnitude of response
YLQPRTFLL	S	4 donors	12.220
VYSTGSNVF	S	1	3.755
VYSSANNCTF	S	1	2.250
VYDPLQPEL	S	1	3.500
TLDSKTQSL	S	2	3.250
RLFRKSNLK	S	1	2.669
QYIKWPWYI	S	1	6.273
QYIKWPWYI	S	1	2.980
LYNSASFSTF	S	1	4.662
KSNLKPFER	S	1	3.062
CYFPLQSYGF	S	1	2.633
CVADYSVLY	S	2	2.000
AEIRASANL	S	1	4.667
KAYNVTQAF	N	1	28.749
YANRNRFLY	M	1	5.556
SSDNIALLV	M	1	2.250
SELVIGAVIL	M	1	2.125
RIAGHHLGR	M	1	3.375
FAYANRNRF	M	1	9.750
EELKKLLEQW	M	1	2.299
AVILRGHLR	M	1	3.062
ATSRTLSYY	M	1	5.318
FLAFVVFLL	E	1	16.222

- *Membrane $_{37}$FAYNRNRF$_{45}$ epitope*

This epitope belongs to the M protein and is highly immunogenic, we find that a donor-restricted to HLA-C*12:03 had a magnitude of response of 9.8-fold over the background. However, there are no other studies that show immunogenicity in unexposed, convalescent, or immunized individuals.

3.2.2 Sequence homology of SARS-CoV-2 epitopes to common human Coronaviruses

Reactivity to SARS-CoV-2 structural epitopes was detected in 75% of pre-pandemic samples, 9 donors recognized at least one peptide. Only a third recognized more than 4 epitopes while most only recognized one or two epitopes. Most of the recognized epitopes are restricted to HLA class I A alleles with 14 positive responses, followed by C with 7 responses, and then allele B with 4. Most of the responses were associated with HLA-A*24:02 (n=4) and C*05:01 (n=4) followed by A*02:01.

Sequence homology between SARS-CoV-2 and other human coronavirus (HCoV) was assessed using multiple alignments with Clustal Omega [120] while visualizing sequence identify and biochemical conservation with JalView [101]. We find that recognized peptides have a median of 2 residues conserved between beta (OC-43, HKU-1) and alpha (NL-63, 229E) common cold coronaviruses. We find that about 45% (n=10) of epitopes have 4 or more conserved residues, and the majority (12/22) share less than 3 amino acids with either alpha or beta HCoVs. Looking at the biochemical conservation, the results are similar, but the conservation occurs at critical anchor residues that may be preserving recognition by promiscuous TCRs. For instance, $_{20}$FLAFVVFLL$_{28}$ a peptide from the E protein, is conserved biochemically at anchor residues P2, P7, and P9 with OC-43 and HKU-1, with a reactivity of 16-fold over background, 8 times as high as the average number of SFU for other peptides. In contrast, epitopes like $_{159}$VYSSANNCTF$_{167}$ have 100%

conservation of their C terminal residues (P6-P10) with the alpha HCoV NL-63. Of the 22 reactive peptides, only 2 epitopes had the highest homology to alpha and beta HCoVs, $_{269}$YLQPRTFLL$_{277}$ and $_{1208}$QYIKWPWYI$_{1216}$ both are from the spike protein but only $_{1208}$QYIKWPWYI$_{1216}$ maps to a region of the spike protein that is less susceptible to residue changes seen in VOC, specifically it is found at the C-terminal end of the S2 region of spike between the last heptapeptide repeat and the transmembrane (TM) domain. On the other hand, the YLQ residue is found at the N-terminal domain of S1 neighboring the RBD, a region highly susceptible to mutations and under strong evolutionary pressure to evade humoral immunity.

3.2.3 Sero-reactivity to human Coronaviruses

More than two-thirds of the donors in this pre-pandemic cohort, recognize CD8$^+$ T cell epitopes from SARS-CoV-2 even though at the time of sample collection, the virus was not circulating the world. Therefore, we hypothesized that those individuals with higher reactivity and percentage of epitope recognition were previously exposed to non-pathogenic HCoVs like OC-43 and HKU-1 which may be reflected in higher sero-reactivity to structural proteins like the nucleocapsid. We probed the presence of polyclonal antibodies against the N protein of seven different HCoVs, using a semi-quantitative Rapid Antigenic Protein in Situ Display ELISA (RAPID-ELISA). We used an in vitro system to produce the N proteins of beta HCoVs OC-43, HKU-1, MERS, SARS-CoV, and SARS-CoV-2 and 2 alpha HCoVs, NL63 and 229E. The presence of antibodies against the N protein was determined in triplicates by chemiluminescence and measured as relative light units (RLUs). Donors were grouped according to the number of CD8$^+$ T cell epitopes recognized, and the relative seroreactivity was plotted per HCoV (Fig. 21). Since seroreactivity to MERS was consistently lower than 2.5 RLU, we chose the highest RLU value for MERS samples plus 3 standard deviations as the cutoff to determine reactivity (threshold

RLU=3.7) which is between 4 to 45 times lower than measurements for other HCoVs. Using this threshold we ensure that 80% of the samples are considered seronegative for SARS-CoV-2 and SARS-CoV N protein and that 100% of samples are negative for MERS, a lethal coronavirus endemic to the Middle East that has not circulated in the United States where the samples were collected [121].

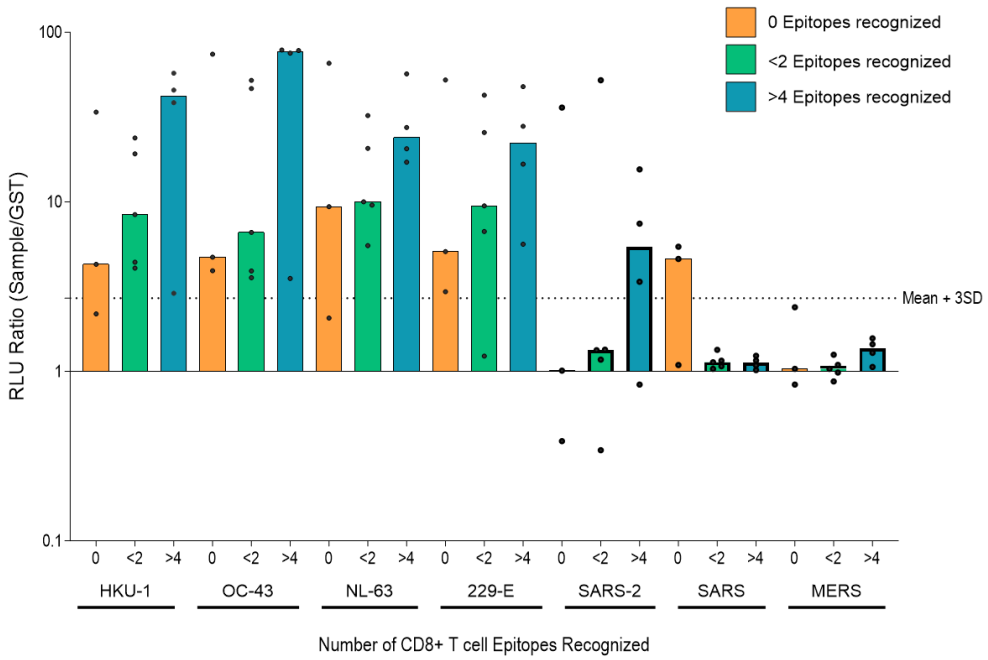


Figure 21. Sero-reactivity to N protein according to the number of CD8⁺ T cell epitopes recognized per donor. Differences are not significant (Kruskal-Wallis test, P>0.999) due to the high variability of outliers and small sample size. RLU ratio higher than 3.7-fold over background is considered sero-reactive for N protein.

Differences in sero-reactivity levels are not significantly different among HCoVs or within the CD8⁺ T cell reactivity groups. However, there is a trend where the median seroreactivity of samples that recognized more than 4 epitopes is higher than those that recognized less than two epitopes, and the same trend is observed for all non-pathogenic HCoVs. All samples were sero-reactive to the N protein of common cold coronaviruses (CCCoV) like HKU-1, OC-43, NL63, and 229-E even though the N protein shares <36%

homology among alpha and beta HCoVs (Fig. 22). Specifically, median seroreactivity to HKU-1 and OC-43 is about 10 and 2 times higher, respectively, for donors that recognize >4 epitopes than for donors that recognize less than 1 peptide.

3.3 Discussion

Understanding how pre-existing SARS-CoV-2-reactive T cells influence susceptibility to COVID-19 is important for defining correlates of protection, finding effective targets for next-gen vaccines, and delineating models of immunity and protection after infection and vaccination. In the context of SARS-CoV-2, T cell reactivity in unexposed individuals is highest for $CD4^+$ T cells at up to 50%, and lower for $CD8^+$ T Cells at around 20%[74, 110, 122, 123]. Surprisingly, we find that 75% of individuals in our pre-pandemic cohort recognize at least one SARS-CoV-2 epitope, and a third recognized more than 4 epitopes from different structural proteins of SARS-CoV-2. This percentage of positive unexposed donors is 3 times higher than previously reported [74, 122-124]. In general T cell pre-existing immunity is higher for $CD4^+$ T cells, up to 50%, likely due to higher peptide promiscuity, or the ability of a peptide to bind different HLA class II alleles. Several studies have found that 50% of $CD4^+$ T cell responses are dominated by promiscuous peptides [125]. In contrast, $CD8^+$ T cell responses are more specific, owing to the closed peptide binding groove of HLA class I alleles that limit the size and biochemical character of the accommodated peptide [126]. A possible explanation for the high rate of positive responses could be related to the long-term in vitro exposure of PBMCs to the tested peptide which may induce expansion of naive T cells [113, 127]. However, the frequency of naive T cells in circulating peripheral blood is extremely low, about 0.03 to 0.21 cells per 10^6 PBMCs/mL. To make sure that responses to peptide stimulation are not due to naive T cells, future experiments could deplete these cells using magnetic beads that remove $CD45RA^+$ $CD8^+$T cells.

On the other hand, reactivity to $CD8^+$ T cell epitopes in unexposed donors may be due to the cross-reactivity of memory T cells (T_{mem}) that originated during prior exposure to a CCCoV. However, homology analysis of the studied peptides revealed that most peptides (20/22) had less than 67% homology which is equivalent to less than 7 residues conserved in 9 or 10 mer epitopes. Still, biochemical conservation was maintained at anchor residues or C terminal ends for 11/22 immunogenic peptides. To further elucidate if reactive peptides are recognized by naive to T_{mem} cells, further stimulations should include phenotyping IFN-γ expressing $CD8^+$ T cells using cell surface markers like CD45RO, CD45RA, CD28, and CCR7.

Human CoVs are responsible for the common cold circulating in the human population and are responsible for symptomatic re-infections every 1-3 years as exemplified by the high seroreactivity to at least three of the four CCCoVs. A report in 2010 on the prevalence of antibodies against HCoVs in a military veteran cohort found that 100% of individuals had antibodies against the N protein of OC-43, 91% to HKU-1, 98% to NL-63, and 99% to 229-E [128]. It is logical to assume that pre-existing responses to SARS-CoV-2 in pre-pandemic samples are due to this cross-reactivity. However, sequence homology between structural proteins from beta HCoVs (HKU-1 and OC-43) and SARS-CoV-2 is less than 36% (depending on the protein) and even less for alpha HCoVs (NL-63, 229-E) (Appendix E, F). On the other hand, patients who recovered from SARS-CoV have T_{mem} cells that are reactive to the N protein 17 years after being infected during the 2003 epidemic in Asia [110]. These T_{mem} cells are cross-reactive to peptides from the N protein of SARS-CoV-2, the homology between the spike proteins of these two viruses is about 73.4%, 89.7% for N and 78% for M, more than twice as much as between SARS-CoV-2 and structural proteins from OC-43 (<35%) and HKU-1 (<35%) (Fig. 22). Reactivity to 15-20 mer epitopes from the N protein was dominated by $CD4^+$ T_{mem} cells and one of the peptides

tested, included the CD8+ T cell epitope characterized in this study, $_{263}$KAYNVTQAF$_{271}$. Therefore, it is possible that the 28-fold reactivity to $_{263}$KAYNVTQAF$_{271}$ is due to CD4+ T cells, however, the 10-day in vitro stimulation done prior to the 24-hour ELISpot plate incubation included IL-2 and IL-7 cytokine which should bias expansion to CD8+ T cell populations [129].

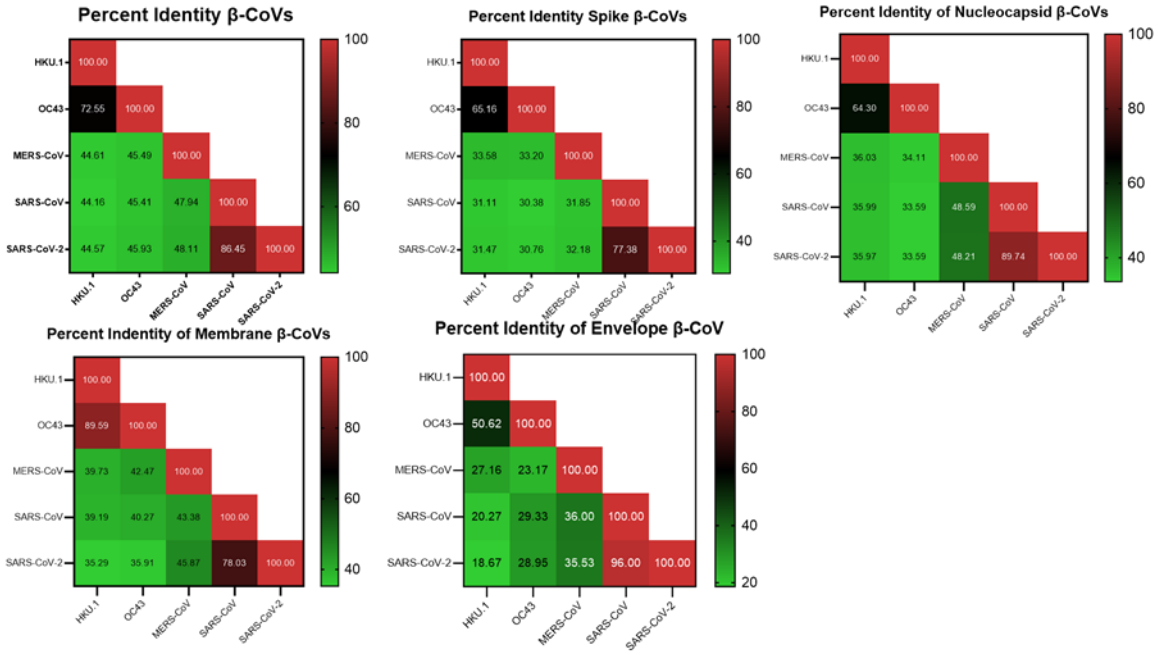


Figure 22. Heat map showing the sequence conservation among β-HCoVs. Panel A compares the complete proteome of SARS-CoV2 against HKU-1, OC-43, MERS-CoV, and SARS-CoV. B, Spike. C, Nucleocapsid. D, Membrane. E, Envelope.

Here we show that most individuals have sero-reactivity to all HCoVs except, as expected, to pathogenic SARS-CoV and MERS-CoV. We show that some unexposed individuals can recognize at least one structural SARS-CoV-2 epitope possibly due to cross-reactive TCRs. A study by Kundu et al shows that T_{mem} cells specific to one of the epitopes described in this study, $_{1208}$QYIKWPWYI$_{1216}$, is part of a protective cross-reactive response against SARS-CoV-2. $_{1208}$QYIKWPWYI$_{1216}$-specific CD8+ Tmem cells associated with protection against SARS-CoV-2 infection among individuals who were exposed to the virus but did not have a positive PCR test or symptoms associated with COVID-19 disease

[57]. Further characterization of immunogenic epitopes in unexposed individuals may provide information about the influence of pre-existing immunity on the severity of COVID-19. Individuals with pre-existing memory T cells could start an immune response faster and prevent severe disease. In fact, (Show study of CD8 T cells mediating asymptomatic infections).

3.4 Limitations and future directions

We did not phenotype SARS-CoV-2 reactive $CD8^+$ T cells, further experiments should include functional assays like ICS or cytokine screens and isolation of single $CD8^+$ T cells sorted by memory phenotype markers to isolate, characterize, clone, and express cross-reactive TCRs like those reactive to the $_{1208}$QYIKWPWYI$_{1216}$ epitope. In addition, non-structural proteins that have higher sequence identity to common cold HCoVs and thus a higher level of cross-reactivity should be tested and identified. We have defined a set of 20 immunogenic SARS-CoV-2 $CD8^+$ T cell epitopes that are not highly conserved between HCoVs and SARs-CoV-2 but are reactive in unexposed individuals. This may be explained by 1) a higher level of cross-reactivity than expected that is generated from exposure from other animal coronaviruses (e.g. dogs, cats), 2) experimental error that leads to activation of naive T cells, or 3) that at a 30-40% level of conservation, TCR reactivity and recruitment of public TCRs is efficiently mounting immune responses that are not specific but may be protective when encountering a novel virus.

3.5 Materials and Methods

We used ELISpot IFN-γ secretion assays to probe the $CD8^+$ T cell reactivity to predicted epitopes and determined sero-reactivity to different HCoV N proteins using RAPID-ELISA. We also performed epitope sequence alignments between HCoVs using Clustal and JalView applications. For a detailed description of these methodologies please refer to Chapter 2.

Chapter 4

Pre-existing Cross-reactive Immunity to SARS-CoV-2

4.1 Introduction

Cytotoxic $CD8^+$ T cells are an essential element in the process of viral clearance. $CD8^+$ T cells recognize viral peptides presented on the cell surface of infected cells like a little flag being hosted up for recognition by the Human Leukocyte Antigen class I (HLA-I) molecule. This interaction is mediated by the T cell receptor (TCR), a heterodimeric molecule composed of an α and β chain, each with a constant and a variable region that confers specificity to the pMHC complex [130]. Each variable chain is generated through a process of somatic gene rearrangement that involves nucleotide insertions and deletions from a group of four regions: variable (V), and joining (J) for the α chain, plus an additional diversity (D) region for the β chain with a constant (C) gene fragment for both chains (Fig. 23) [131]. This recombination process confers high variability to the TCR and is multiplied by the different α and β pairings, which in theory could generate up to 10^{18} different αβ TCR pairs [132]. However, personal diversity is estimated to be less than 10^8 TCR pairs due in part to the negative selection (central tolerance) that occurs during T cell development in the thymus, which eliminates T cells with a high affinity for self-pMHC [132]. T cell diversity is also limited by the carrying capacity of the individual where it is estimated that there are less than 10^8 naive TCRs and a maximum of 10^{12} T cells in a human [133]. The diversity of the TCR is based on six complementarity-determining regions (CDRs), three from the α variable region and 3 from the β, that engage with the peptide: MHC complex [130]. Still, this high diversity is not enough to be specific to all possible pathogens, calculated to be over 10^{15}, not taking into account post-translation modifications that generate novel p:MHC complexes [134]. As a result, T cell cross-

reactivity caused by TCR binding degeneracy may help to respond to the immense number of p:MHC complexes generated during a viral infection or tumorigenesis.

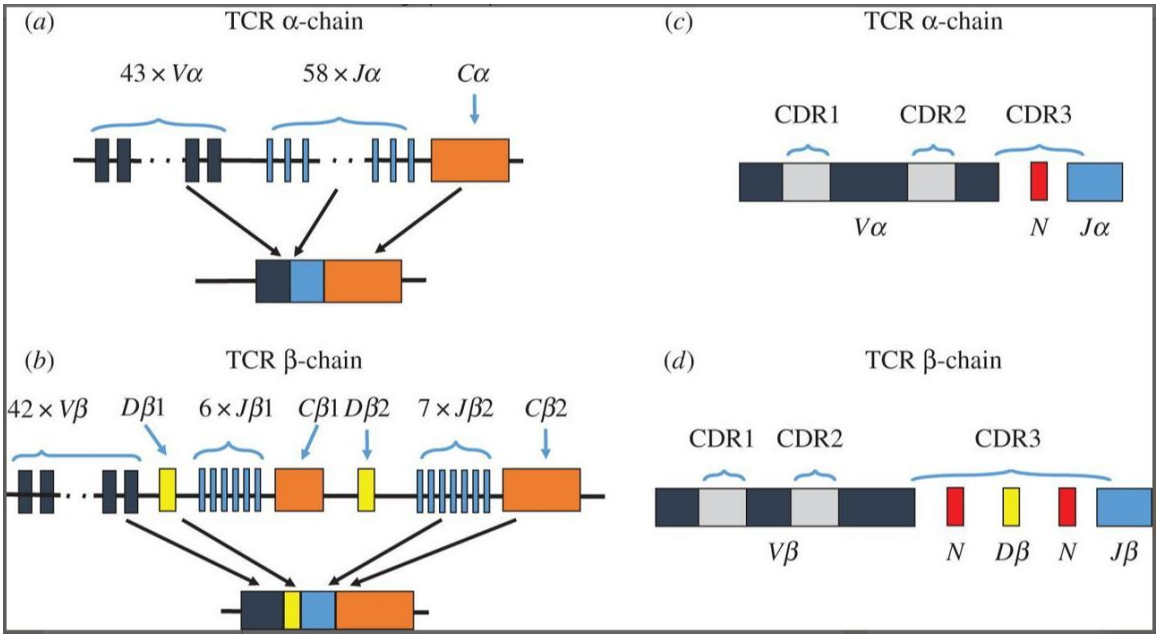


Figure 23. T-cell receptor rearrangement. The top left panel represents the Variable (V), Joining (J), and Constant (C) regions of the α chain. The bottom panel shows the V, J, and C regions plus a Diversity (D) region of the β chain. Semi-stochastic recombination, nucleotide deletions, and addition generate the final TCR αβ pairs. Adapted from ref [135].

Pre-existing immunity to SARS-CoV-2 has been characterized in 20-50% of unexposed individuals likely due to prior exposures to common cold coronaviruses (CCCoVs). Understanding the impact of pre-existing responses and cross-reactivity at the individual and population levels may provide insight into the differential COVID-19 outcomes that range from asymptomatic to severe and lethal. It can also influence epidemiological models of herd immunity and affect the performance of vaccines, thus affecting vaccination regimens and public health management protocols. Accordingly, the role of pre-existing SARS-CoV-2 cross-reactivity will remain a central topic of research. A growing number of spike and non-spike T cell responses have been identified, specifically against non-structural proteins and the N structural protein in unexposed donors [57, 110,

122]. One study identified non-spike cross-reactive $CD8^+$ T cells that are associated with protection from infection in exposed individuals. Specifically, IL-2-only secreting $CD8^+$ T cells detected 1-6 days after exposure with SARS-CoV-2 were found after stimulation with HKU-1, OC-43, and SARS-CoV-2 peptide pools in individuals that were PCR negative and not in PCR^+ or symptomatic individuals [57].

Here, we isolated cross-reactive T cells generated after stimulation with two highly immunogenic epitopes from the N (KAYNVTQAF or KAY^N) and M (FAYANRNRF or FAY^M) proteins in an unexposed donor. After sequencing, we expressed the reactive TCRs using an $\alpha\beta^{-/-}$ Jurkat T cell line, J76-NFAT-Luc-RE (Nuclear Factor of Activated T cells), that upon activation of the TCR signaling pathway will express luciferase which can be used as a semi-quantitative-readout for T cell activation [136]. After expression, the engineered T cell lines will be tested for cross-reactivity to KAY^N and FAY^M peptides that are homologous from common cold human Coronaviruses (HCoVs) (Fig. 24).

4.2. Results

4.2.1 A pre-pandemic donor is highly reactive to SARS-CoV-2 epitopes from the Nucleocapsid and Membrane structural proteins

To find reactivity to SARS-CoV-2 in pre-pandemic samples, as described in Chapter 3, we tested predicted $CD8^+$ T cell peptides using IFN-γ ELISpot secretion and Activation Induced Marker assays as a readout of reactivity. One of the 12 pre-pandemic donors, BCP15, a 59 y/o white male was highly immunogenic to non-spike structural epitopes. Specifically, we found reactivity to the nucleocapsid (N) peptide KAYNVTQAF (KAY^N) and the membrane (M) peptide FAYNRNRF (FAY^M). KAY^N increased reactivity 28-fold over background as measured by IFN-γ ELISpot and expression of activation markers like CD137 was 52 times higher than AIM expression in unstimulated cells or the negative control (Fig. 25). The KAY^N peptide was predicted by EnsembleMHC (EnMHC) to be

restricted to HLA-A*32:01, HLA-B*15:03/ B*15:17/ B*46:01 /B*57:01 and HLA-C*12:03. Previous reports show that the KAY^{N} peptide can be presented by HLA-B*57:01 but has also been reported by other algorithms (MHC-Flurry and NetMHCpan 4) to be restricted to HLA-C*12:03. Since our donor is restricted to HLA-C*12:03 and no other possible HLAs, we infer that the KAY^{N} response in this donor is likely restricted to HLA-C*12:03. The second epitope FAY^{M} is also restricted to HLA-C*12:03 and had a 10-fold increase in reactivity as measured by IFN-γ ELISpot and 6 times higher level of AIMs (Fig. 25).

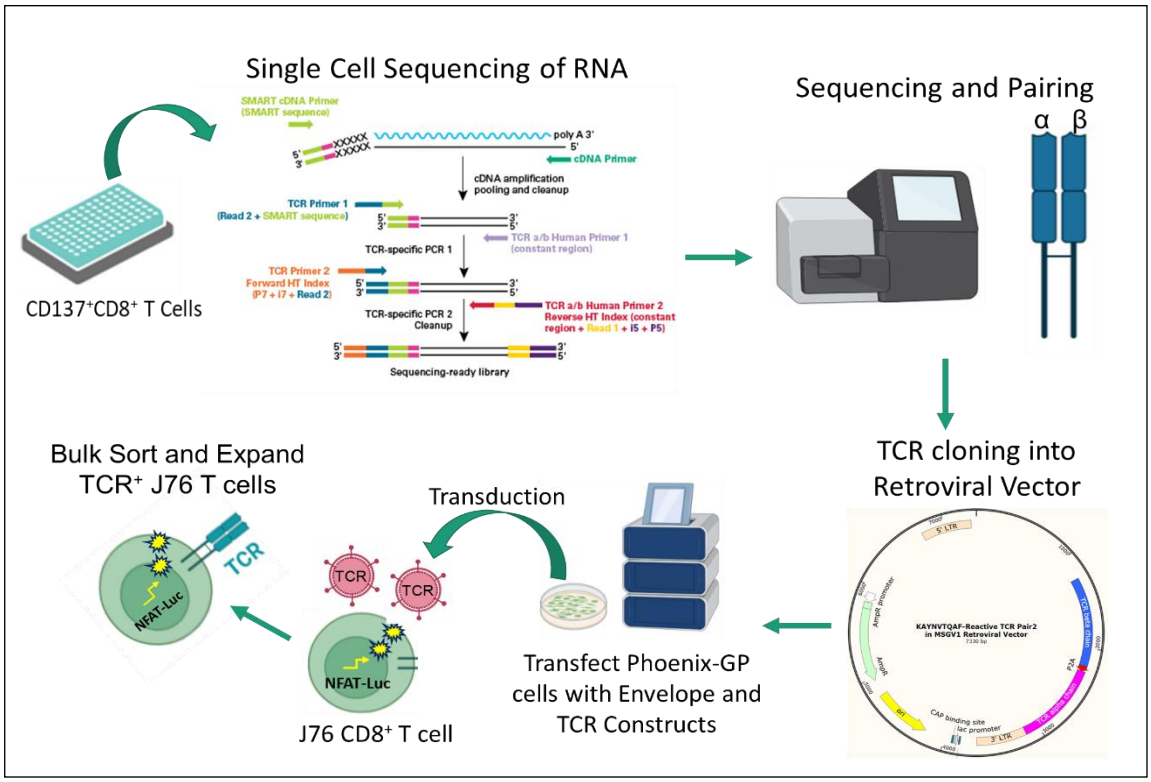


Figure 24. Workflow to isolate, sequence, clone and express SARS-CoV-2-reactive TCRs from a pre-pandemic donor. PBMCs isolated from pre-pandemic donor BCP15 were stimulated with KAY^{N}-pulsed autologous APCs. CD137^{+} CD8^{+} T cells were single-cell sorted into a 96-well plate and processed. After sequencing, KAY^{N}-reactive α/β TCR pairs were identified and cloned into the MSGV1 and pLenti-CMV-Puro vectors. mTCR+ cells were bulk sorted and expanded to be tested for reactivity and specificity to the KAY^{N}:C*12:02 complex.

4.2.2 KAY^{N} stimulated PBMCs express 5 different α/β novel pairs.

To isolate KAY^{N} and FAY^{M}-reactive CD8^{+} TCRs, we single-cell sorted KAYN-reactive CD8+ T cells sorted the top brightest CD8^{+} T cells (top 30%) into a 96 well plate pre-filled with lysis buffer and RNAse inhibitor for an RT-PCR reaction using barcoded primers from the SMARTer® Human scTCR a/b profiling kit. After RT-PCR, the cDNA sequence was amplified and verified for size and quality/quantity (concentration of DNA) for further processing, using the Agilent Tapestation (Appendix G). We found that most pooled cDNA libraries (12 pooled samples) were the correct size (~250-750 bp) and had enough DNA to be used for further processing. After 2 consecutive PCR reactions and a final tapestation analysis (Appendix H), samples were sequenced and verified against the TCRβ data from the ImmunoSeq T cell receptor database (see methods) and the IMGT/V Quest and VDJdb online analysis tools. After verification, we found a total of 20 different α/β TCR pairs, and after alignment using the IgBlast tool (ncbi.nlm.nih.gov/igblast/igblast.cgi, see methods) we found 11 pairs with productive (not truncated or out of frame) α and β TCR pairs but of those, 2 pairs were reported as part of Jurkat repertoire, 4 were previously reported to be associated to general immune responses to respiratory viruses and only 5 were novel pairs with 1-α and 2-β chains not previously reported (Table 13, Appendix I). The TCRα and TCRβ chains were linked by the furin 2A self-cleaving peptide (P2A) and cloned into the MSGV1 retroviral vector using the NcoI and EcoRI restriction sites (Synthesized by Genscript, NJ. US) (Fig. 30, Appendix J). To clone into the pLenti-CMV-Puro vector, the TCR α/β genes were PCR amplified from the MSGV1 construct as a template, with primers that added flanking attB1/attB2 regions for use with the Gateway cloning system into the pLenti-CMV-puro vec tor (Fig. 30-Methods section 4.4).

Table 13. TCR α and β genes and CDR3 α and β variable sequences. Sequences in bold are novel pairs. Genes marked with a star have not been reported in the literature (VDJ or NCBI). Controls 1 and 2 are Jurkat TCRs isolated from positive controls.

Pair	TCRα genes	TCRβ genes	CDR3a	CDR3b
Control 1	TRAV10/TRAJ20	TRBV12-4/TRBD1/TRBJ1-4	CVVSYPTNDYKLSF	CASSLKDSEKLFF
Control 2	TRAV12-1/TRAJ6	TRBV12-4/TRBD1/TRBJ1-4	CVVSASSGGSYIPTF	CASSLKDSEKLFF
1	**TRAV22/TRAJ49**	**TRBV7-2/TRBJ1-1 ★**	**CAVERGTGNQFYF**	**CASSLLISRFEAFF**
2	**TRAV21/TRAJ58**	**TRBV7-2/TRBJ1-1 ★**	**CAGETSGSRLTF**	**CASSLLISRFEAFF**
3	TRAV19/TRAJ49	TRBV27/TRBD2/TRBJ2-5	CALSPGGNQFYF	CASSPGTTQETQYF
4	**★TRAV19/TRAJ49**	**TRBV12-3/TRBD1/TRBJ1-2**	**CALSPGGNQFYF**	**CASSFSTCSANYGYTF**
5	TRAV19/TRAJ49	TRBV20-1/TRBJ1-5	CALSPGGNQFYF	CSASEGWGNQPQHF

4.2.3 KAY-reactive TCRs are successfully expressed in Jurkat J76-NFAT-Luc-RE CD8⁺ T Cells

Expression of the 5 different KAY-reactive TCRs was unsuccessful when using the retroviral system MSGV1, we experimented with retronectin-coated plates to increase transduction efficiency and treating the viral supernatant with polybrene as previously described [137]. Transduction efficiencies were never over 11% for mTCR⁺ J76 cells (Fig. 26). Nevertheless, we sorted mTCR⁺ T cell lines for pairs 1 and 2 and expanded for three days to make stock for future experiments at a purity of ~92% mTCR⁺ cells (4.3.4). Also, the MSGV1 vector does not allow for the selection of positive clones, so after sorting the TCR expression started to decline after the fourth passage. As a result, the α/β pairs were cloned into the pLenti-CMV-Puro vector (kind gift from Dr. Rahman, ASU). Transduction efficiencies increased to 50-98% depending on the TCR pair (Fig. 26, Appendix K).

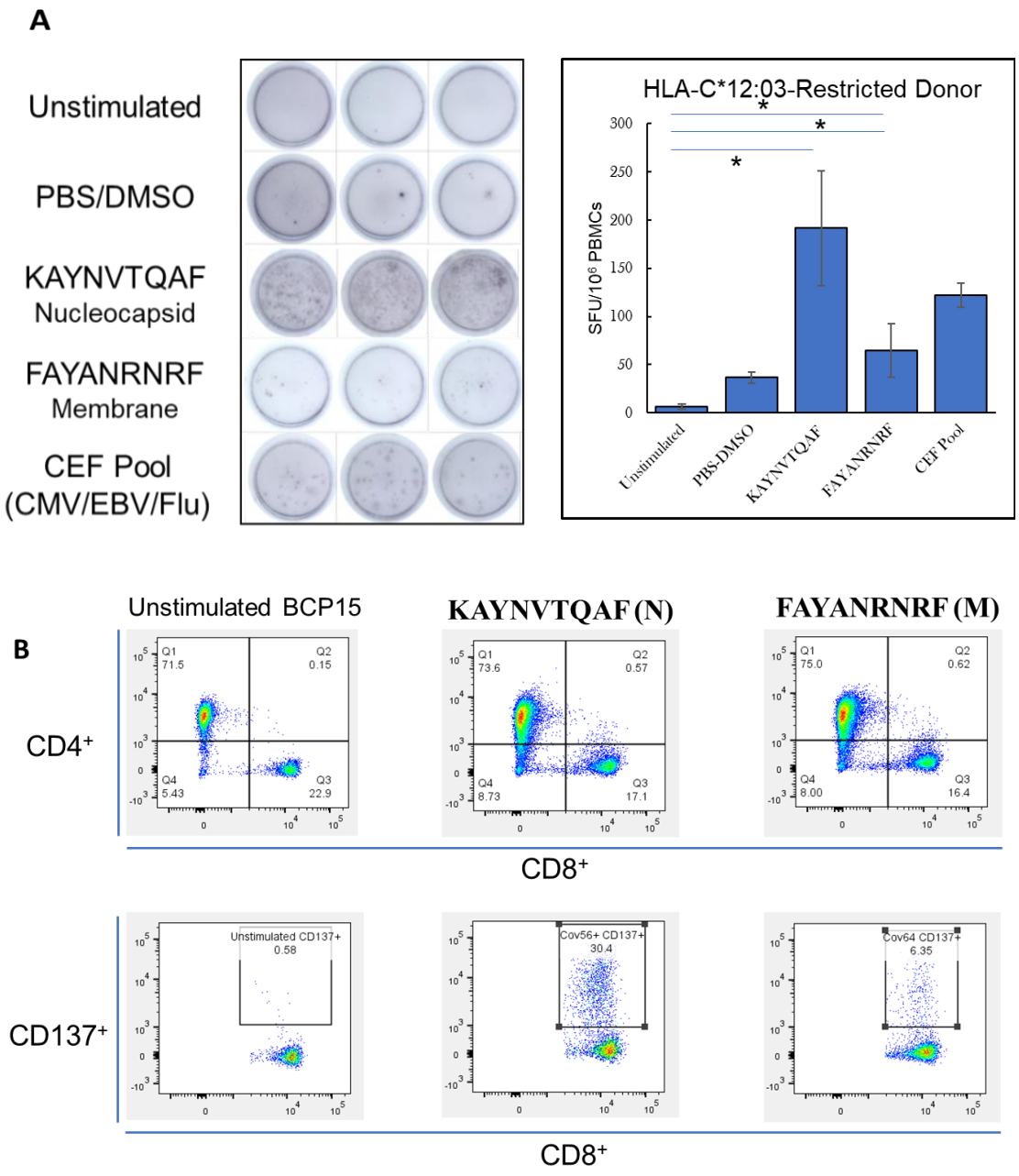


Figure 25. Structural epitopes KAYNTVQAF^{N} and FAYANRNRF^{M} are highly reactive in a pre-pandemic donor. Peptide pulsed autologous APCs were co-incubated with PBMCs isolated from donor BCP15. Experiments were done in triplicates. A) ELISpot IFN-γ secretion; negative controls: unstimulated and PBS/DMSO wells; positive control: CEF peptide pool. Right panel: histogram of reactivity by spot forming units (SFU) per 10^6 PBMCs. B) AIM assay, flow cytometry panel gated for CD4$^+$, CD8$^+$ and CD137$^+$ T cells. CD137$^+$ CD8$^+$ T cells were single-cell sorted for further processing.

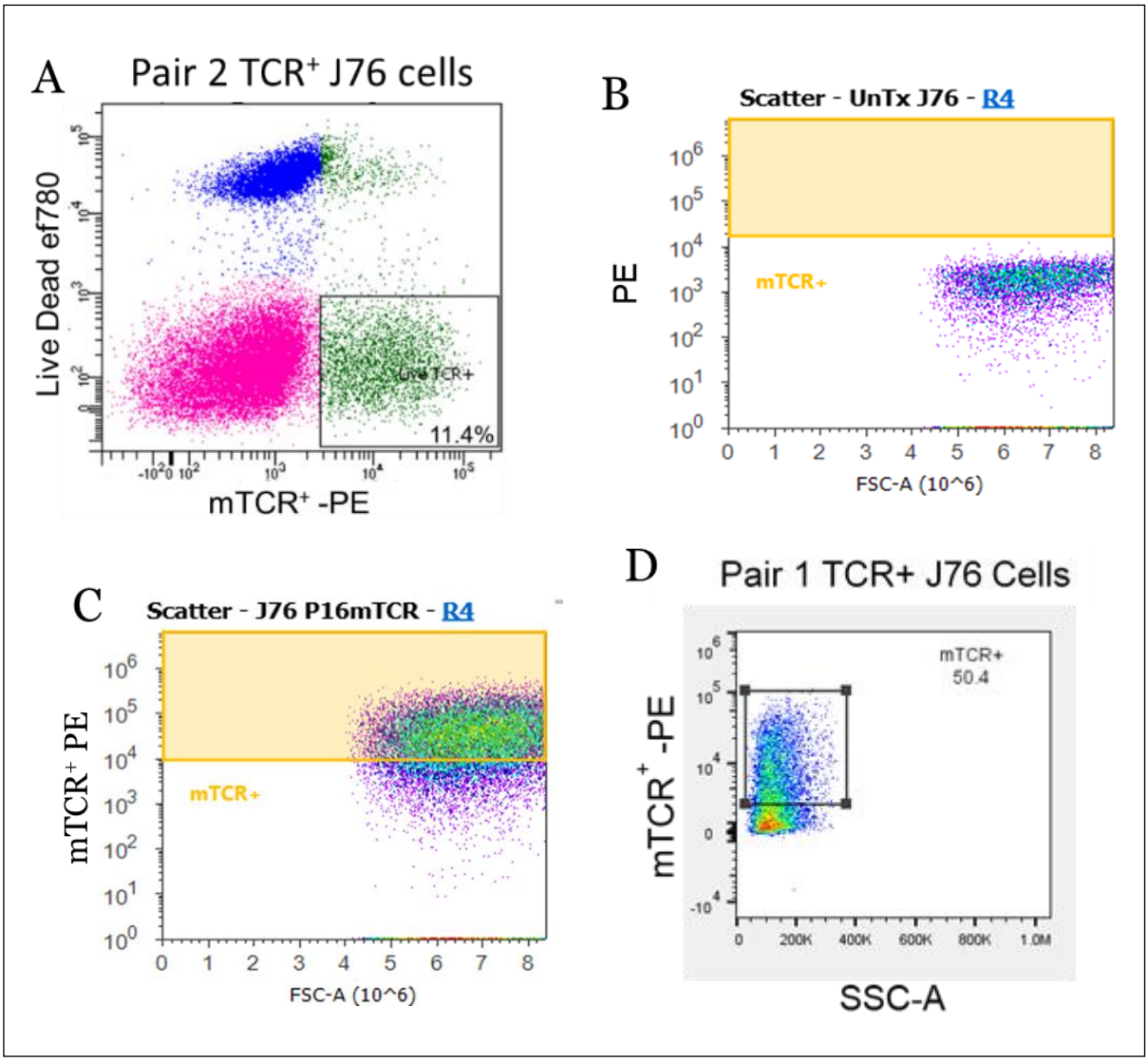


Figure 26. A. Bulk sort of mTCR+ J76-CD8+ transduced cells with retroviral supernatant expressing the KAY^{N}-reactive TCR Pair #2 (MSGV1 vector). Cells were sorted by staining with anti-mTCR-PE and gating for viability with zombie violet. B) Untransduced J76 CD8+ T cells, mTCR-, C) Bulk sorted Pair#2 TCR J76 Cells after 3 expansions, D) mTCR+ expression after transduction with the pLenti-CMV-Puro Pair 1 TCR construct. PE: Phycoerithrin fluorophore; SSC-A: side scatter-Area, ef780: live dead stain.

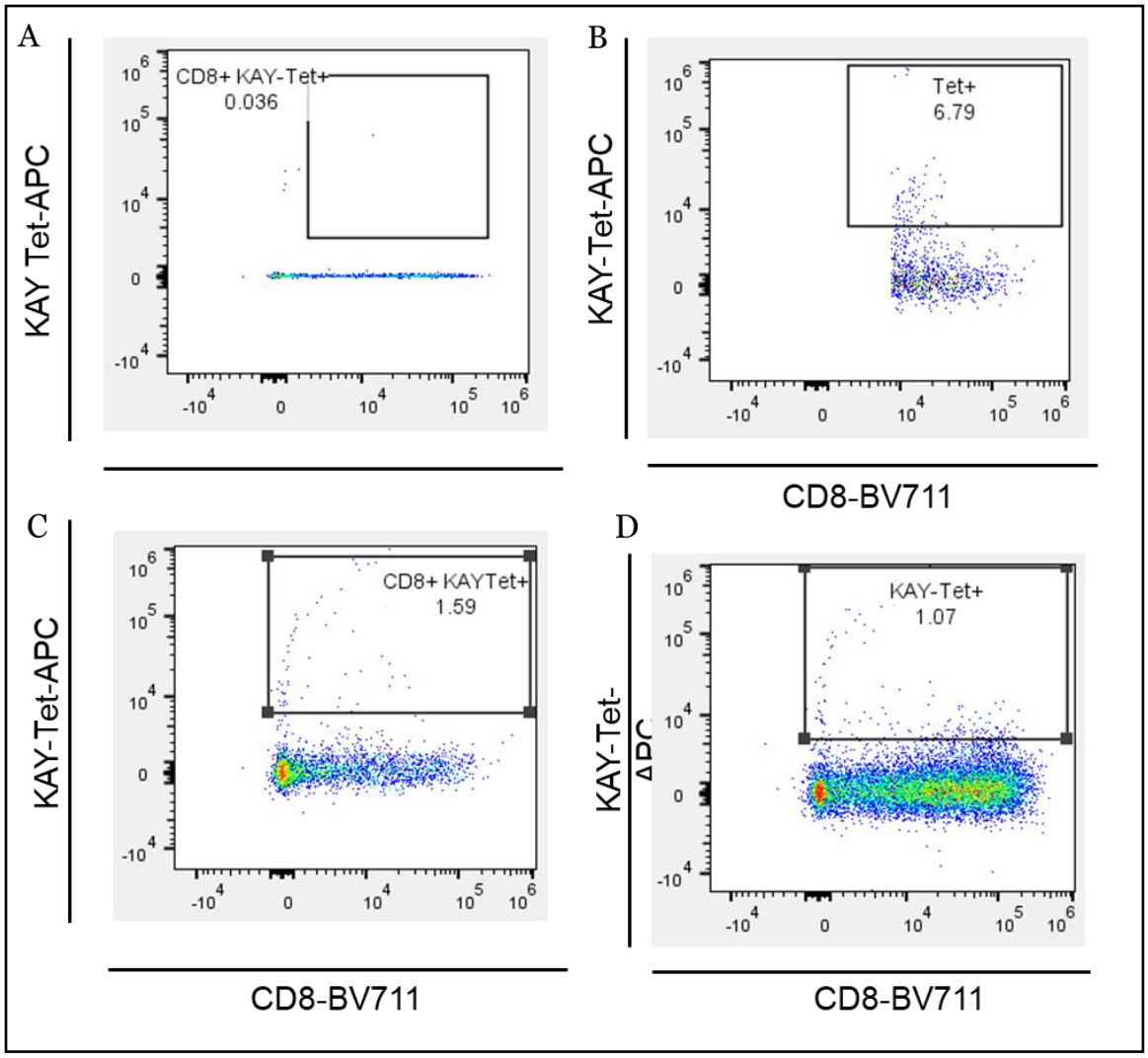


Figure 27. KAY^{N}-Tetramer-PE stain of Pair 1 (A), 2 (B), 3 (C), and 5 (D) of KAY^{N}-reactive TCRs in MSGV1. KAY^{N}-TCR T cell lines (J76-NFAT-Luc-RE). Flow cytometry plots after bulk sorting and expansion of mTCR^{+} T cell lines. Live cells (ef780-) were gated on CD8+ versus KAY^{N}-tetramer (PE).

4.2.4 KAY-reactive TCR cells may not be KAY-specific

We produced 2 T cell lines that express pair 1 and 2 (MSGV1 retroviral vector) and stained with a KAY-HLA-C*12:03-PE tetramer (NIH Tetramer Core Facility). The mTCR^{+} CD8^{+} T cells were not stained by the KAY-tetramer. We then tried to stimulate PBMCs derived from the same donor with the KAY^{N} peptide for 10 days and again, we were unable

to detect any KAY+ T cells with the tetramer (Fig. 27). We tested the newly made Lentiviral KAY-reactive T cell lines but the specificity was not improved given that the mTCR+ expression level increased by more than 5 fold (Appendix K). It is possible that the tetramer is not stable enough to stain at room temperature as a pink precipitate formed after staining that was separate from the cell pellet. Also, the tetramer was not freshly prepared and may have been degraded at the time of testing. The specificity of PBMCs stimulated with peptide-pulsed autologous APCs was not evaluated.

4.2.5 K562 cells express HLA-A*02:01, B*57:01 and C*12:03

We transfected 293LX cells, with the pLenti-CMV-Puro vectors expressing either HLA-A*02:01, HLA-B*57:01, HLA-C*12:03 to be used as artificial antigen-presenting cells. The main purpose of these cell lines will be to present peptides to the KAY-reactive TCR cell lines. Each cell line was sorted and expanded under puromycin selection (0.75ng/mL) for 15 days, aliquoted in FBS with 10% DMSO at a density of 2 x 10^6 cells /mL, and stored in liquid nitrogen until needed (Appendix L).

4.3 Discussion

Sequence conservation of KAY^N and FAY^M epitopes in comparison to other HCoVs is 55.56% and 44.5%, respectively (Appendix F). Although neither epitope has a high identity to any of the CCCoVs, it shares the highest similarity to OC-43 and HKU-1. The KAY^M epitope shares higher biochemical conservation where P1, P4, P5, P6, P7, and P9 maintain their biochemical identity for 67% of the residues. The FAY^M epitope shares homology to both OC-43 and HKU-1 (~45%) but shares 78% biochemical characteristics at residues P1, P3-P7, and P9. In addition to the high biochemical conservation, the binding motifs for HLA-C*12:03 represented by the logo plots in Figure 28, are a match for P2, P3, P7, and P9, all reported to be very important for higher binding affinity which is linked to immunogenicity [138]. MHC-Flurry predicts a binding affinity for KAY^M and C*12:03 of

29.9 nM, with a probability of being processed of 0.75 and presented by HLA-C*12:03 of 0.97567. The FAY^{M} epitope also has a predicted high binding affinity of 38.8 nM, with a lower probability of presentation (0.346) but a high probability of being presented by C*12:03 of 0.88. Reactivity to both epitopes is consistent with current literature where about a third of CD8^{+} T cell responses are generated against N- and M- derived peptide pools that include both epitopes and also elicit CD4^{+} reactivity of HLA class II alleles [90]. We found 5 different α/β novel pairs isolated from KAY-reactive CD8^{+} T cells (Table 13). None of the pairs have been reported in the literature to be associated with CD8^{+} T cell responses against a virus or a malignant cell. However, only three single chains, 2α and 1β, were completely novel. The other 3α and 4β chains were reported to be associated with virally infected cells (HIV, SARS-CoV-2) or tumor antigens (Glioma, Merkel cell polyomavirus, Human papillomavirus). After successful expression of 2 pairs (pair 1 and 2) of KAY^{N} reactive TCRs in J76 CD8^{+} NFAT-Luc-RE Jurkat cells (J76-CD8^{+}), we bulk sorted CD8^{+} and mTCR^{+} (constant region of the mouse TCR) cells and rested them overnight.

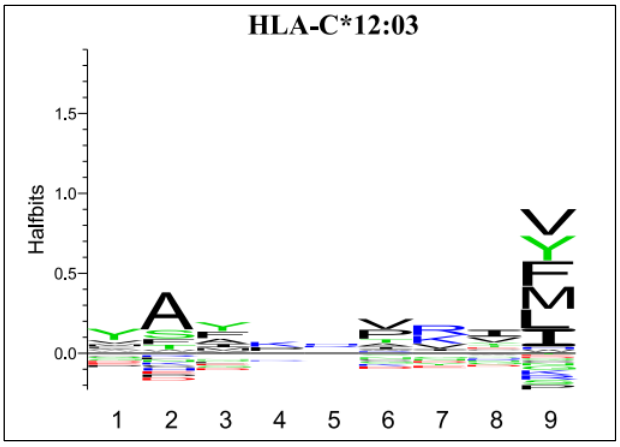


Figure 28. Logo plot of the peptide binding motifs of HLA-C*12:03. Generated with the Seq2Logo server. The height of each letter indicates the impact on binding at the specific position. Amino acids on the positive side of the y-axis have a positive effect on binding to the HLA allele, while the ones on the negative side have the opposite effect [139].

Each cell line was expanded three times, from 96 well plates at a density of about 5 x 10^4 cells per well, to 2 x 10^5 cells per 24 well. Population numbers were amplified to about 16 x10^6 cells and half of the cells were aliquoted to about 1 million cells per mL and frozen in case of contamination during further expansion. The other half was expanded but the frequency of $mTCR^+$ cells started to decrease from 96% after sorting, to about 45% after the second expansion, down to 8% after the third expansion. Since we could not select the cells with antibiotics because the retroviral vector does not include an antibiotic-selective marker, we decided to clone the TCR pairs into a lentiviral vector. pLenti-CMV-Puro does not require stimulation of PBMCs with anti-CD3 (OKT3) before transduction, it requires polybrene in solution but eliminates the need for retronectin-coated plates and has a Puromycin selective gene. We performed an antibiotic curve to select the appropriate amount of Puromycin to use, and after 7 days, 50% of untransduced J76 cells were dead at a concentration of 1.4 ug/mL of Puromycin, so our working concentration was 0.7 ng/mL of media. With this setup, we increased transduction efficiency from 11% or less to 65-87% for pairs 3, 4, and 5. Pairs 1 and 2 had lower transduction efficiencies (32-38%) and as a result, will not be prioritized for specificity or functional testing. Future experiments will test the specificity of all 5 pairs to KAY^M-pulsed antigen-presenting cells (APCs) generated from dendritic cells, autologous APCs from donor BCP15, or artificial APCs restricted to HLA-C*12:03.

About 12 different TCRβ CDR3 sequences have been identified for the KAY^N peptide and about 18 for the FAY^M epitope, both associated with individuals that were either exposed and uninfected or had asymptomatic COVID-19, but from different HLA restrictions [82, 90]. Since staining with tetramer was unsuccessful, it is possible that the monomer generated is unstable and misfolds during tetramerization or that partial unfolding decreases avidity to the TCR. It is also possible that expression of the TCRs on

the cell surface is not high enough to be detected by tetramer stain, however, staining in parallel with a subset of the T cell lines with anti-mTCR-PE, shows that at the moment of tetramer staining, between 70 to 87% of the cells were $mTCR^+$. Since we did not test the lentiviral generated T cell lines with KAY^N-tetramer, it is not possible to know if improved expression prior to sorting, reflects proper folding and α/β dimerization of the TCR.

It is also possible that the identified TCRs have low binding affinity for the KAY^N:C*12:03 complex because they are part of a subset of effector responses characterized by low-binding affinity to the target, a common occurrence during an immune response against viral pathogens and cancer [140]. Specific to SARS-CoV-2, a few studies have characterized low avidity and broad-cross reactivity of memory $CD4^+$ T cells in unexposed individuals [141, 142]. Further, in vivo, responses to infection show that weak TCR:pMHC interactions can activate a naive $CD8^+$ T cell response and be sufficient to generate effector and memory $CD8^+$ T cells [143]. Thus, it is possible that the isolated TCRs do not have high binding affinity to the KAY^N-tetramer, however, we can determine if the TCR recognizes KAY^N:C*12:03 complexes by challenging the lentiviral-generated T cell lines with decreasing concentrations of peptide-pulsed (0.5, 0.25, 0.1, 0.0025, 0.001 mg/mL) artificial APCs (aAPC) or dendritic cells (DCs) and quantify the levels of activation by measuring luciferase expression or by ELISpot IFN-γ secretion ELISpot assay.

In addition to levels of antigen and TCR expression, CD8 co-receptor expression seems to be critical for T cell activation, especially for low-affinity interactions [144]. Thus, I tested the levels of CD8 expression in the $\alpha\beta^{-/-}$ J76-$CD8^+$ T cell line before transduction and found that our stocks were about 40% $CD8^+$. As a result, the J76-$CD8^+$ Jurkat cells were selected with hygromycin for one week, the $CD8^+$ T cell expression increased to 85% before transduction. The generated T cell lines were bulk sorted and placed in liquid N for later experimentation with the expectation that tetramer staining would be improved.

To end, the TCR sequences linked to both epitopes are unique and are not associated with public TCR responses (TCR clonotypes that are shared by more than one individual) which indicates that the TCRs isolated after KAY^N stimulation may be specific and be part of the cross-reactive repertoire against SARS-CoV-2 in unexposed individuals. Further experiments should include testing the lentiviral T cell lines for KAY^N-tetramer staining, challenged with either DCs or aAPC. In addition, IFN-γ secretion assays and AIM assays should be performed to detect activation of PBMCs from the same donor by homologous KAY^N and FAY^M peptides from OC-43, HKU-1, NL-63, 229E, SARS-CoV and MERS (as additional negative control).

4.5 Materials and Methods

First, we isolated peripheral blood mononuclear cells (PBMCs) from a heathy unexposed pre-pandemic donor, as summarized in Chapter 2 (see Materials and Methods section). The donor is a 59 y/o white male. During PBMC isolation, about one million cells were aliquoted to isolate RNA for HLA typing using exome sequencing. Once HLA typed, we tested the reactivity of predicted peptides according to HLA restriction.

4.5.1 SARS-CoV-2 ancestral peptide synthesis

Epitope prediction and selection were done using EnsembleMHC as previously described (Chapter 2-materials and methods) and two epitopes with high reactivity were selected for further study: KAYNVTQAF for the N protein and FAYNRNRF from the M protein (ProImmune, UK. >80% purity). We used a CEF pool (CMV, EBV, and Flu viral epitopes) (ProImmune, UK) as a positive control and three negative controls, untreated cells, PBS with 5% DMSO, and with the human cytomegalovirus peptide pp65 (hCMVpp65) NLVPMVATV that is restricted to the HLA-A*02:01.

4.5.2 Autologous APC generation from healthy individual PBMCs:

Autologous CD40L-activated B cell APCs were generated by incubating healthy donor PBMCs (peripheral blood mononuclear cells) with irradiated (32Gy) K562-cells expressing CD40L (KCD40L) at a ratio of 4:1 (800,000 PBMCs to 200,000 irradiated KCD40Ls). Cells were kept in a 24-well plate at 37°C and 5% CO_2 in B cell media (BCM) consisting of IMDM (Gibco), 10% heat-inactivated human serum (GemCell 100-512), 100 mM HEPES (Gibco) and 2 mM L-Glutamine (Gibco). BCM is supplemented with 10 ng/mL of recombinant human IL-4 (R&D Systems), 2 μg/mL Cyclosporin A (Sigma Aldrich), and insulin transferrin supplement (ITES, Lonza). On day 7, APCs were washed with PBS, expanded into a 24-well plate, and re-stimulated with fresh irradiated KCD40Ls. We determined APC purity by CD19+/CD86+ antibody staining and quantitation by flow cytometry (CD19+/CD86+ T cells >90% of total PBMCs). APCs can be re-stimulated for up to 5 weeks or cryopreserved until needed for new expansions.

4.5.3 T cell stimulation by autologous APCs:

Healthy donor B-cell APCs were stimulated with individual peptides and cultured in BCM complete and recombinant IL-4 (Gibco-51500-056). BCM complete is made of IMDM (Life technologies-12440061), 5% human serum (GemCell-100-512), 1X L-Glutamine (Gibco025030-081), 1X HEPES (Gibco-15630-080), 1X antimycotic-antibiotic (Gibco-15140-122). After 24 hours, the APCs were incubated with whole PBMCs at a ratio of 1:3.65 (400,000 APCs: 1,400,000 PBMCs) in a 24-well plate with BCM, 20 U/mL recombinant human IL-2 (rhIL-2 Fisher Sci.-202IL05-CF) and 5 ng/mL IL-7 (Fisher Sci.-207IL005CF). A partial media exchange was done on day 5 and freshly peptide-pulsed APCs were added. Each reactive well was split in half to perform ELISpot and AIM assays.

4.5.4 ELISpot assay

Enzyme-linked immunosorbent assay (ELISpot) was performed as previously described (Chapter 2) [102]. Each peptide and control were tested in triplicates. Samples that did not have more than 5 SFU/10^6PBMCs for the positive control were not considered for analysis.

4.5.5 Flow cytometry staining for activated-T cells (AIM).

Stimulated PBMCs were stained for activation markers, specifically CD137 for $CD8^+$ T cells and CD134 for CD4+ T cell markers to quantify epitope-specific T cell activation [103]. We used the Attune™ NxT Flow cytometer in conjunction with the FlowJo™ software for data acquisition and analysis.

4.5.6 Single cell sorting to isolate the TCR of $CD8^+$-SARS-CoV-2 reactive T cells:

After stimulation of PBMCs by autologous APCs with the immunogenic epitopes activated T cells were single-cell sorted into a 96-well plate previously loaded with lysis buffer and RNAse inhibitor mix and processed according to the TAKARA kit instructions (TAKARA 634431). Single-cell sort was done using the CD137-PE marker using the BD FACSAria™ Illu cell sorter (BD Biosciences #643245). Data was analyzed using FlowJo™ software.

4.5.7 SMARTer® Human scTCR a/b profiling

TCR α/β libraries were prepared according to the manufacturer's instructions. Briefly, first-strand DNA synthesis is done using a proprietary reverse primer and a set of SMART-Seq Indexed oligos so that each sequenced TCR α and β chain can be traced back to a well by indexing to a row (A-H) 6-base identifier or barcode (Fig. 29). In-plate negative (n=4) and positive (n=4) controls are used per 96 well plates, they consist of no cells (negative control) and a control Jurkat RNA. The eight control wells are placed as

indicated in Figure 29. This is followed by a cDNA amplification step to increase the amount of the template required for 2 subsequent PCR reactions. The cDNAs are then pooled by column so that each pool will contain cDNA from 8 different wells. The cDNA is purified by using AMPure XP beads to remove primers and primer dimers. Pooled and purified cDNA is validated and quantified using the Agilent 2100 Bioanalyzer and High sensitive DNA kit (Agilent #5067-4626). Two TCR-specific PCR reactions follow that will selectively amplify the TCR sequences from the cDNA generated. The first will add an Illumina sequence at the 5' end of the cDNA and a TCRa/b reverse primer at the 3' end of the constant region. The second PCR will amplify the obtained the full-length sequences obtained in the previous PCR and add an Illumina HT sequencing adapter. DNA is then purified with AMPure XP beads and the library is quantified and validated again using the Agilent 2100 Bioanalyzer. At this point, libraries are pooled to a final concentration of 4nM. Pooled libraries are diluted to a final concentration of 13.5 pM and spiked with 5-10% PhiX Control (Illumina FC-110-3001).

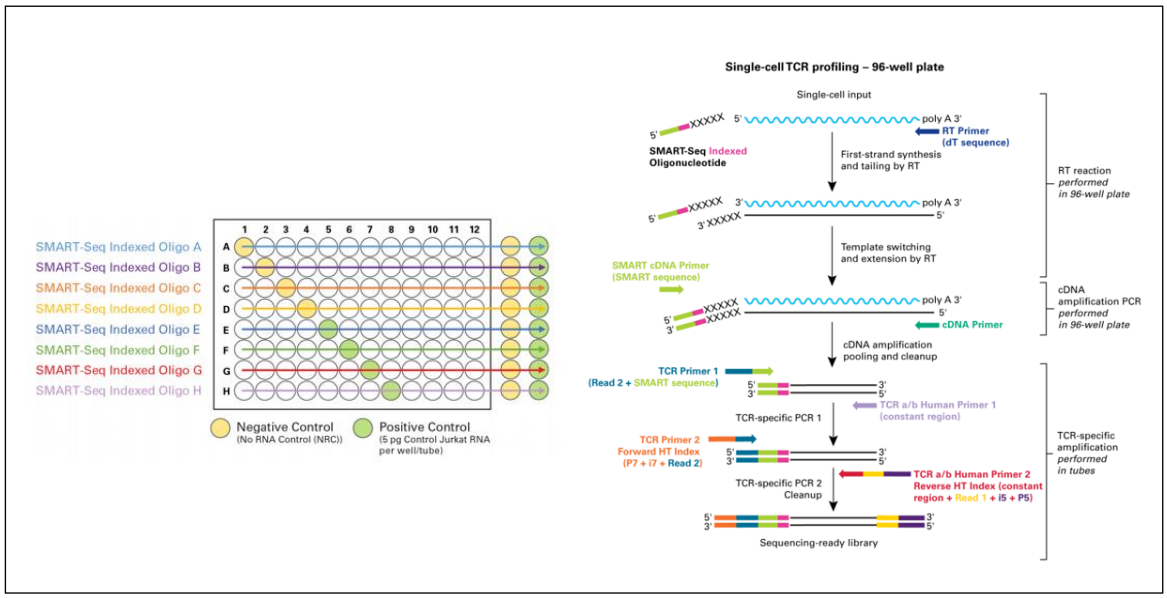


Figure 29. Single-cell sequencing of peptide-reactive TCRs. A) Plate map of barcode assignment and placement of positive and negative control wells per plate. B) Single-cell TCR processing workflow. From Takarabio.com user manual (Cat#634431).

4.5.8. Library sequencing and data analysis

Libraries were sequenced on Illumina MiSeq v.3 instruments (2x300bp paired-end reads) using the 600-cycle MiSeq Reagent Kit v3 (Illumina # MS102-3003) at the Genomics Core Facility at Arizona State University (ASU). Subsequent raw sequence analysis used MiXCR (v.3.0.7) and VDJ tools (v.1.2.1) at the Bioinformatics Core (ASU). The retrieved TCRβ sequences were analyzed against existing TCRβ data from the ImmunoSeq TCR database (https://clients.adaptivebiotech.com/immuneaccess) and confirmed using the IMGT/V Quest and VDJdb online analysis tools. To make sure that the pairs were productive, an alignment was performed using the IgBlast tool (ncbi.nlm.nih.gov/igblast/igblast.cgi) [145].

4.5.9 Cloning of productive TCRs in retroviral (MSGV1) and lentiviral (pLenti-CMV-puro) vectors

Only pairs with productive α and β TCRs were synthesized and subcloned into the MSGV1 retrovirus expression vector (GenScript Biotech NJ, USA) [146] as previously described [147]. The TCRα and TCRβ chains were linked by the furin 2A self-cleaving peptide (P2A) and cloned into the MSGV1 vector using the NcoI and EcoRI restriction sites (Fig. 30). To prevent mispairing with endogenous TCR (if using PBMCs), the human TCR constant region was substituted with the mouse TCR (mTCR) constant regions. The MSGV1-BFP (Blue fluorescent protein) was used as a positive control for transfection efficiency. MSGV1-JM22-TCR (TCR specific for the HLA-A*02:01 restricted FluM1 epitope) and MSGV1-E6-TCR (TCR specific for the HLA-A*02:01-restricted human papillomavirus (HPV) epitope TIHDIILECV from the E6 viral protein were used as positive controls.

Cloning using lentiviral vectors was done to improve transduction efficiency, TCR pairs were cloned into the pLenti-CMV-Puro (Addgene-38481) vectors and GFP-pLenti was used as a positive control for both transfection and transduction efficiency [148]. The TCR α/β pairs were PCR amplified to be cloned into the CMV-pLenti-Puro vector. Five forward primers for each α and 5 each β variable chains, included an attB1 region, a Shine-Delgarno and Kozak regions, and a 5' complementary region to the Variable region of each α or β chain (Appendix J). The reverse primer was the same for all chains, it included a complementary region to the 3'- end of the mouse constant region, followed by two stop codons and an attB2 region.

To prevent mispairing with endogenous TCR (if using PBMCs), the human TCR constant region was substituted with the mouse TCR (mTCR) constant regions. The MSGV1-BFP and MSGV1-JM22-TCR/ MSGV1-E6-TCR constructs were used as positive controls for transfection and transduction efficiency respectively.

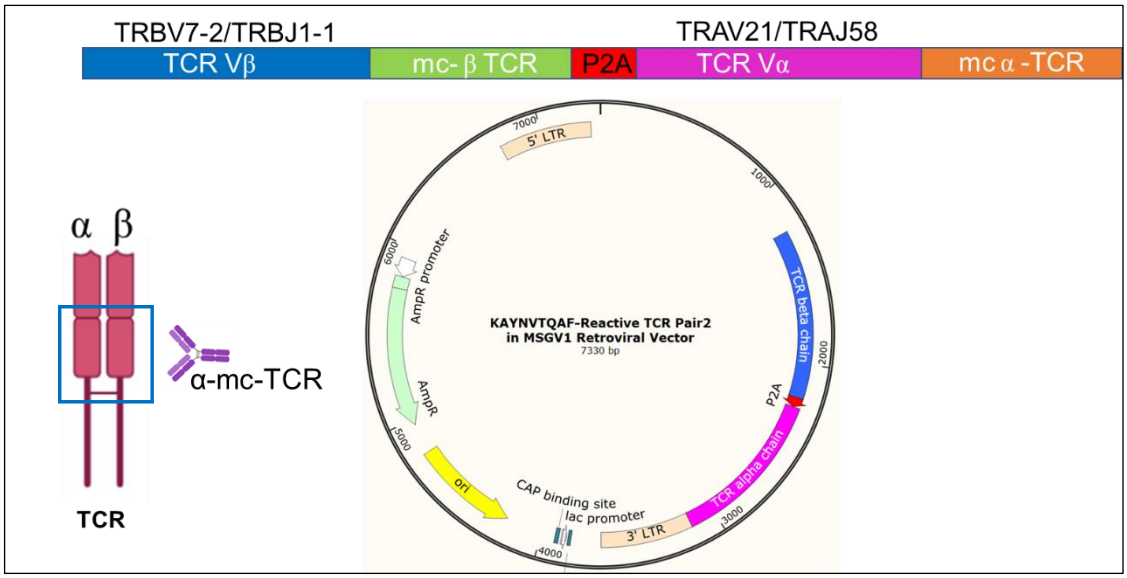


Figure 30. MSGV1 Retroviral constructs with α/β TCR pair. Top: The gene for the β TCR variable region is followed by the mouse constant β region and is joined to the TCR α variable and constant regions by the self-cleaving peptide P2A. The pairs were cloned into the MSGV1 vector using the NcoI and EcoRI sites (Genscript). α-mc-TCR: antibody against the mouse constant region of the TCR. Map made with Snapgene. Figure created with BioRender.

4.5.10 Expression of SARS-CoV-2-reactive TCRs in PBMCs and Jurkat-NFAT-Luc-RE CD8$^+$ T cells

4.5.10.1 Transfection

To make SARS-CoV-2-reactive T cell lines, we followed the procedure established by Draper et al [147]. Briefly, 293GP cells were plated in 6 well plates at a 0.75-0.8 x 10^5 cells/mL 28-20 hours prior to transfection. Cells were co-transfected with MSGV1-TCRα/β pairs and the envelope RD114 using Fugene HD (Promega 72050) at a ratio of 1:6 of Fugene to DNA not exceeding 5 ug per well. Cells were spun at 1000g (~2250 rpm) for 30 minutes and rested overnight. Partial media change was done the next morning and again in the afternoon. Viral supernatant is collected 48- and 72 hours post-transfection. Viral supernatant is centrifuged at 300 g for 6 minutes to clear the viral supernatant from the cellular pellet, aliquoted into 2 mL tubes, and stored until needed at -80°C.

4.5.10.2 Transduction

a. *Stimulation of PBMCs:* Cells were plated with TCM (T cell media: RPMI, 10% human serum, antibiotics and antimycotic), IL-2 (300 IU/mL), and OKT3 (50ng/mL) 24 hours before transduction.
b. *Coating of plates with viral particles:* 6 well non-tissue culture plates were coated with 1.5 mL of retronectin solution (20ug/mL in PBS) overnight at 4°C, and blocked prior to viral incubation with 2 mL of 2% BSA (Bovine serum albumin) in PBS (phosphate buffered saline) for 30 minutes at room temperature. Viral supernatant was added to the plates (after discarding the blocking solution) and centrifuged for 2 hours at 2000 g at 32°C. Discard the viral supernatant and add the stimulated PBMCs at a concentration of 0.5-1 x10^6 cells/well. Spin for 10 minutes at 1500 rpm, acceleration/brake set at 1. Leave in the incubator overnight

and change the media in the morning. Check GFP or BFP expression 72 hours post-transduction.

c. *Test for successful TCR transduction:* Cells were stained with anti-CD8-BV711, mTCR-PE, and Zombie Violet for viability. To test for tetramer binding we stained with anti-CD8-BV711, and HLA-C*12:03 monomer, tetramerized with streptavidin-PE. GFP (for Lentiviral constructs) and BFP (for retroviral constructs) were also stained for viability, and GFP^+ or BFP^+ cells were quantified as a control for successful transfection/transduction, respectively.

4.5.11 Cell culture

293GP cells are cultured in DMEM, 10% FBS with no antibiotics before transfection. Then, half of the media was replaced with antibiotic-free media and replaced after overnight incubation with a transfection mix.

4.5.12 Quantification and Statistical Analysis

Data was analyzed using Microsoft Excel and all statistical analysis was performed with GraphPad Prism 10 (GraphPad).

CHAPTER 5

SUMMARY AND CONCLUSION

The current study validates the use of a consensus predictive algorithm, EnMHC, to find predicted peptides with high binding affinity, and a high probability of being cleaved and presented on the cell surface. Most importantly, about 83% of tested ligands are highly immunogenic, which underscores the use of consensus platforms to increase the probability of finding true epitopes. We provide evidence that the precision of predictive algorithms may be higher than expected. Of the 17 non-immunogenic peptides, 14 were immunogenic in other studies, most under the same HLAs as the ones predicted by EnMHC or with different restrictions that were also predicted by EnMHC (but not tested due to the HLA restriction of the cohort) or MHC Flurry. Even under EnMHC restrictive thresholds, we find that the binding capacity of most individuals may be enough to protect against severe disease. For instance, the donor with the lowest binding capacity (ASC-95) of 5 structural proteins and 23 nsp reported mild covid-19 symptoms, which indicates that the binding capacity of most individuals may be enough for an efficient $CD8^+$ T cell response. For most donors, binding capacity ranges from 5 to 23 different structural peptides and can span up to 134 peptides from the entire proteome. Our cohort is too small to find any correlations between binding capacity and protection or risk associated with COVID-19 and we lack an acute or severe COVID-19 disease cohort to make comparisons. However, we validated in vitro the immunogenicity of 76 SARS-CoV-2 $CD8^+$ T cell epitopes, including 27 structural epitopes not reported elsewhere and 29 that were part of unconvoluted peptide pools. In total, we show the individual immunogenicity of 56 peptides and confirm/expand the reported HLA restriction of 75 epitopes.

After individual screening on a cohort of 19 individuals restricted to 31 different HLA class I alleles, we found a set of 13 immunodominant, highly conserved epitopes from all

structural proteins. The HLA class I and epitope combinations are predicted to cover about 91.15% of the North American population and about 93.5% of the global population. Their use in a multiepitope string of pearl type of vaccine will supplement the use of spike as the main immunogen and increase the breadth and resilience of the vaccine to new VOC. Of the 13 immunodominant epitopes, 7 are immunoprevalent as they are predicted to be presented by at least 3 different HLA class I alleles. Specifically, 11/13 epitopes are reactive 12 months post-vaccination in individuals who self-reported as never infected and 7 of those 11 epitopes are from non-spike structural proteins. We hypothesize that the presence of $CD8^+$ reactive T cells to non-structural proteins may be the result of antigenic spread, also known as an antigen cascade, where the presence of one immunogen enhances the response to neighboring immune cells with different antigenic specificities. The only possible explanation for uninfected individuals being reactive to non-spike structural proteins is that they were exposed to SARS-CoV-2 and were asymptomatic for COVID-19. An extensive body of work provides evidence of the existence of $CD8^+$ T_{mem} cells in individuals who were exposed to SARS-CoV-2 but were not PCR positive, sero-reactive to S or RBD nor had COVID-19 symptoms [1, 74, 110, 122]. Although we did not phenotype the epitope reactive $CD8^+$ T cells, it is highly likely that recognition of immunodominant epitopes in samples from uninfected donors after vaccination is mediated by T_{mem} cell populations generated by asymptomatic infection or continuous exposure to the virus. Given that all self-reported uninfected donors are part of the ASU academic community, their level of exposure to the virus during the months of sample collection for the 6 months (June-ASC-12/ASC-25; August-ASC7, October-ASC-63 of 2021) and 12-month (September-ASC-12/ASC-25 of 2021 and February-ASC-7; April-ASC-63 of 2022) time points was high, about 21% positivity rate on average during the fall of 2021, more than

2000 weekly cases per 100,000 Arizona residents during Feb 2022 and the highest rates of infection for individuals between 19 and 24 years old during the same periods [149].

Here, we provide evidence that peptides from the neighboring regions of RBD and the S2 region of spike are highly conserved among VOC and other HCoVs, and that immunogenicity to these epitopes is sustained over time. Since we did not phenotype the $CD8^+$ SARS-CoV-2-reactive T cells, we are not able to determine if the sustained immunogenicity is coming from T_{mem} populations generated from previous infections or vaccination. However, the present data shows that vaccination after infection, increases reactivity to spike and most importantly, to other structural epitopes with the potential to do the same to other non-structural proteins. The implication is that vaccination not only enhances responses to the main immunogen, but in populations where most individuals have been infected, the use of booster shots may enhance immunogenicity to structural and non-structural proteins. Comparison between pre-pandemic reactivity and post-pandemic ASU samples indicate an increase in the breadth and magnitude of the $CD8^+$ T cell response after vaccination and provides evidence that memory against spike in previously infected donors, sustains T_{mem} cells in circulation with different structural specificities. Most vaccination efficacy studies test the immunogenicity of spike epitopes and do not consider immunogenicity to other SARS-CoV-2 proteins. Future studies on the efficacy of booster vaccinations should include conserved M, N, and E targets to determine the level of $CD8^+$ T cell reactivity after boosting and assess if maintained responses are protective against severe COVID-19 disease.

Of the 76 reactive peptides, 10 are mutated in VOC and most of those amino acid changes are present in the latest Omicron subvariants and affect epitopes close to or within the RBD. We show that reactivity to variant peptides is maintained or increased in 9/10 variant epitopes which suggests that $CD8^+$ T cell responses against new variants may be

sustained especially in vaccinated individuals. Reactivity in convalescent unvaccinated individuals was overall higher than in pre-pandemic samples but lower than in convalescent vaccinated donors. A study of T cell reactivity changes to VOC epitopes found that $CD4^+$ and $CD8^+$ T_{mem} cell responses to Omicron subvariants induced by prior infections or BNT162b2 (Pfizer-BioNTec) vaccination are reactive to epitopes from the B.1.1.529 variant [150]. The same study found that the magnitude of $CD4^+$ and $CD8^+$ T cell responses was lower in convalescent individuals but not in those who were also vaccinated with the Pfizer model [150]. Our data provides evidence that vaccination enhances T cell responses that supplement the waning antibody immunity and enhance protection against new VOC and that T cell recognition is not impaired in most $CD8^+$ T cell conserved targets, even with one or two amino acid changes.

About 59% of pre-pandemic donors are sero-reactive to the SARS-CoV-2 N protein and all of them have antibodies to OC-43 which could explain the high $CD8^+$ T cell reactivity to SARS-CoV-2 structural epitopes (75%). Given the high variability of conservation between HCoVs and SARS-CoV-2, it is likely that cross-reactivity is centered on specific regions of structural proteins. For instance, none of the pre-pandemic donors are reactive to the N protein of pathogenic coronaviruses like MERS and SARS. Sequence alignment of the N proteins shows that a section of the C-terminal domain of the protein is highly conserved between SARS-CoV-1/2 and MERS but is not shared with the other alpha and beta CCCoVs. Given the high level of sero-reactivity to OC-43, NL63, and 229E N protein, it is possible that infection with CCCoVs enhances $CD8^+$ T cell reactivity to SARS-CoV-2 structural proteins much like the vaccine does in those that are convalescent. The possibility of enhanced $CD8^+$ and $CD4^+$ T cell reactivity to SARS-CoV-2 epitopes from booster vaccinations as well as infection with common cold CoV should be evaluated, using

conserved regions of structural and especially non-structural proteins to determine the extent of protection, if any, from cross-reactive interactions.

To end, we isolated 5 pairs of KAYNVTQAF^{N} reactive TCRs and successfully engineered 5 different T cell lines using a lentiviral system that allowed for selection of the stable cell lines with puromycin. Although the TCRs were not specific to the KAY^{N} peptide, it is possible that the isolated receptors may be specific to the homologous peptides from HCoVs. It is also possible that the isolated TCRs have low binding affinity for the KAY^{N}:C*12:03 complex characteristic of a general immune response against viral pathogens in unexposed individuals. Characterization of other SARS-CoV-2 TCRs isolated from spike-reactive CD4^{+} T_{mem} cells in naive donors have low binding affinity and high cross-reactivity [141, 142]. Further characterization of the engineered T cell lines and their TCRs will provide insight into cross-reactive, naive responses known to be protective against novel pathogens and even VOC.

An effective immune response against SARS-CoV-2 characteristic of individuals with asymptomatic and mild COVID-19 disease integrates an early humoral response (IgG and IgA) with early CD4^{+} and CD8^{+} T cell responses. Most of the uncomplicated infections were characterized by a T_{h1} response [151-153]. Unlike other viral infections (influenza and flavivirus), structural proteins are immunodominant, with spike being the most immunodominant probably due to its large size and high level of expression [90]. Vaccination increases recognition of spike epitopes, but it also sustains T cell populations reactive to M, N, and E proteins and potentially nsps. Prior infections with CCCoVs generate a level of cross-reactivity that may explain why 80% of individuals have uncomplicated SARS-CoV-2 infections and recover from COVID-19. CD8^{+} T cell responses are durable in convalescent individuals up to 6 months post-infection while vaccination elicits long-lasting recognition of up to 12 months in the current study. This is supported

by previous reports that find memory CD4 and CD8 T cells in convalescent individuals up to 8 months post-infection with a small decline in the frequency of those populations [95]. Overall, the breadth and persistence of the T cell responses characterized in this study, indicate that VOC may not completely evade T cell recognition at the population level and underscore the importance of adding additional structural and non-structural targets to the next generation of vaccines.

REFERENCES

1. Rydyznski Moderbacher, C., et al., *Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity*. Cell, 2020. **183**(4): p. 996-1012 e19.
2. Peng, Y., et al., *An immunodominant NP(105-113)-B*07:02 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease*. Nat Immunol, 2022. **23**(1): p. 50-61.
3. Peng, Y., et al., *Broad and strong memory CD4 (+) and CD8 (+) T cells induced by SARS-CoV-2 in UK convalescent COVID-19 patients*. bioRxiv, 2020.
4. Cohen, K.W., et al., *Longitudinal analysis shows durable and broad immune memory after SARS-CoV-2 infection with persisting antibody responses and memory B and T cells*. Cell Rep Med, 2021. **2**(7): p. 100354.
5. Pishesha, N., T.J. Harmand, and H.L. Ploegh, *A guide to antigen processing and presentation*. Nat Rev Immunol, 2022. **22**(12): p. 751-764.
6. Peters, B., et al., *The design and implementation of the immune epitope database and analysis resource*. Immunogenetics, 2005. **57**(5): p. 326-36.
7. Medhasi, S. and N. Chantratita, *Human Leukocyte Antigen (HLA) System: Genetics and Association with Bacterial and Viral Infections*. J Immunol Res, 2022. **2022**: p. 9710376.
8. Sidney, J., B. Peters, and A. Sette, *Epitope prediction and identification- adaptive T cell responses in humans*. Semin Immunol, 2020. **50**: p. 101418.
9. Calis, J.J., R.J. de Boer, and C. Kesmir, *Degenerate T-cell recognition of peptides on MHC molecules creates large holes in the T-cell repertoire*. PLoS Comput Biol, 2012. **8**(3): p. e1002412.
10. Tenzer, S., et al., *Antigen processing influences HIV-specific cytotoxic T lymphocyte immunodominance*. Nat Immunol, 2009. **10**(6): p. 636-46.
11. La Gruta, N.L., et al., *A virus-specific CD8+ T cell immunodominance hierarchy determined by antigen dose and precursor frequencies*. Proc Natl Acad Sci U S A, 2006. **103**(4): p. 994-9.
12. Pang, K.C., et al., *Immunoproteasome subunit deficiencies impact differentially on two immunodominant influenza virus-specific CD8+ T cell responses*. J Immunol, 2006. **177**(11): p. 7680-8.
13. Sette, A., et al., *The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes*. J Immunol, 1994. **153**(12): p. 5586-92.
14. Chen, W., et al., *Immunoproteasomes shape immunodominance hierarchies of antiviral CD8(+) T cells at the levels of T cell repertoire and presentation of viral antigens*. J Exp Med, 2001. **193**(11): p. 1319-26.
15. Crowe, S.R., et al., *Differential antigen presentation regulates the changing patterns of CD8+ T cell immunodominance in primary and secondary influenza virus infections*. J Exp Med, 2003. **198**(3): p. 399-410.

16. Valkenburg, S.A., et al., *Molecular basis for universal HLA-A*0201-restricted CD8+ T-cell immunity against influenza viruses*. Proc Natl Acad Sci U S A, 2016. **113**(16): p. 4440-5.

17. Berman, H.M., et al., *The Protein Data Bank*. Nucleic Acids Res, 2000. **28**(1): p. 235-42.

18. Saper, M.A., P.J. Bjorkman, and D.C. Wiley, *Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 A resolution*. J Mol Biol, 1991. **219**(2): p. 277-319.

19. Sette, A. and J. Sidney, *Nine major HLA class I supertypes account for the vast preponderance of HLA-A and -B polymorphism*. Immunogenetics, 1999. **50**(3-4): p. 201-12.

20. Sidney, J., et al., *HLA class I supertypes: a revised and updated classification*. BMC Immunol, 2008. **9**: p. 1.

21. Nguyen, A.T., C. Szeto, and S. Gras, *The pockets guide to HLA class I molecules*. Biochem Soc Trans, 2021. **49**(5): p. 2319-2331.

22. Calis, J.J., et al., *Properties of MHC class I presented peptides that enhance immunogenicity*. PLoS Comput Biol, 2013. **9**(10): p. e1003266.

23. Chowell, D., et al., *TCR contact residue hydrophobicity is a hallmark of immunogenic CD8+ T cell epitopes*. Proc Natl Acad Sci U S A, 2015. **112**(14): p. E1754-62.

24. Lu, T., et al., *Deep learning-based prediction of the T cell receptor-antigen binding specificity*. Nat Mach Intell, 2021. **3**(10): p. 864-875.

25. Yewdell, J.W. and J.R. Bennink, *Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses*. Annu Rev Immunol, 1999. **17**: p. 51-88.

26. Murata, S., et al., *The immunoproteasome and thymoproteasome: functions, evolution and human disease*. Nat Immunol, 2018. **19**(9): p. 923-931.

27. Kincaid, E.Z., et al., *Mice completely lacking immunoproteasomes show major changes in antigen presentation*. Nat Immunol, 2011. **13**(2): p. 129-35.

28. Oseroff, C., et al., *Dissociation between epitope hierarchy and immunoprevalence in CD8 responses to vaccinia virus western reserve*. J Immunol, 2008. **180**(11): p. 7193-202.

29. Sarkizova, S., et al., *A large peptidome dataset improves HLA class I epitope prediction across most of the human population*. Nat Biotechnol, 2020. **38**(2): p. 199-209.

30. Sidney, J., et al., *Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture*. Curr Protoc Immunol, 2013. **Chapter 18**: p. Unit 18 3.

31. Quach, H.Q., et al., *Evaluating immunogenicity of pathogen-derived T-cell epitopes to design a peptide-based smallpox vaccine*. Sci Rep, 2022. **12**(1): p. 15401.

32. Cheng, Y. and W.H. Prusoff, *Relationship between the inhibition constant (K1) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction*. Biochem Pharmacol, 1973. **22**(23): p. 3099-108.

33. Gulukota, K., et al., *Two complementary methods for predicting peptides binding major histocompatibility complex molecules*. J Mol Biol, 1997. **267**(5): p. 1258-67.

34. Peters, B. and A. Sette, *Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method*. BMC Bioinformatics, 2005. **6**: p. 132.

35. Lundegaard, C., O. Lund, and M. Nielsen, *Prediction of epitopes using neural network based methods*. J Immunol Methods, 2011. **374**(1-2): p. 26-34.

36. Peters, B., M. Nielsen, and A. Sette, *T Cell Epitope Predictions*. Annu Rev Immunol, 2020. **38**: p. 123-145.

37. Reardon, B., et al., *Allele-Specific Thresholds of Eluted Ligands for T-Cell Epitope Prediction*. Mol Cell Proteomics, 2021. **20**: p. 100122.

38. Zhang, C., et al., *Optimally-connected hidden markov models for predicting MHC-binding peptides*. J Bioinform Comput Biol, 2006. **4**(5): p. 959-80.

39. Jurtz, V., et al., *NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data*. J Immunol, 2017. **199**(9): p. 3360-3368.

40. Lundegaard, C., et al., *State of the art and challenges in sequence based T-cell epitope prediction*. Immunome Res, 2010. **6 Suppl 2**(Suppl 2): p. S3.

41. Shao, W., et al., *The SysteMHC Atlas project*. Nucleic Acids Res, 2018. **46**(D1): p. D1237-D1247.

42. Stranzl, T., et al., *NetCTLpan: pan-specific MHC class I pathway epitope predictions*. Immunogenetics, 2010. **62**(6): p. 357-68.

43. Wilson, E.A., et al., *Total Predicted Mhc-I Epitope Load Is Inversely Associated with Population Mortality from Sars-Cov-2*. Cell Rep Med, 2021: p. 100221.

44. O'Donnell, T.J., A. Rubinsteyn, and U. Laserson, *MHCflurry 2.0: Improved Pan-Allele Prediction of MHC Class I-Presented Peptides by Incorporating Antigen Processing*. Cell Syst, 2020. **11**(4): p. 418-419.

45. Andreatta, M. and M. Nielsen, *Gapped sequence alignment using artificial neural networks: application to the MHC class I system*. Bioinformatics, 2016. **32**(4): p. 511-7.

46. Rasmussen, M., et al., *Pan-Specific Prediction of Peptide-MHC Class I Complex Stability, a Correlate of T Cell Immunogenicity*. J Immunol, 2016. **197**(4): p. 1517-24.

47. Zhang, H., O. Lund, and M. Nielsen, *The PickPocket method for predicting binding specificities for receptors based on receptor pocket similarities: application to MHC-peptide binding*. Bioinformatics, 2009. **25**(10): p. 1293-9.

48. Bassani-Sternberg, M., et al., *Deciphering HLA-I motifs across HLA peptidomes improves neo-antigen predictions and identifies allostery regulating HLA specificity*. PLoS Comput Biol, 2017. **13**(8): p. e1005725.

49. Nichols, K., *False discovery rate procedures*, in *Statistical parametric mapping: The analysis of functional brain images*, K.J.F. William D Penny, John T. Ashburner, Stefan J. Kiebel, Thomas E. Nichols, Editor. 2007, Elsevier. p. 246-252.

50. Paul, S., et al., *HLA class I alleles are associated with peptide-binding repertoires of different size, affinity, and immunogenicity*. J Immunol, 2013. **191**(12): p. 5831-9.

51. Johnson, D.R., *Differential expression of human major histocompatibility class I loci: HLA-A, -B, and -C*. Hum Immunol, 2000. **61**(4): p. 389-96.

52. Tarke, A., et al., *Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases*. Cell Rep Med, 2021. **2**(2): p. 100204.

53. Sharma, S. and P.G. Thomas, *The two faces of heterologous immunity: protection or immunopathology*. J Leukoc Biol, 2014. **95**(3): p. 405-16.

54. Welsh, R.M., et al., *Heterologous immunity between viruses*. Immunol Rev, 2010. **235**(1): p. 244-66.

55. Greenbaum, J.A., et al., *Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population*. Proc Natl Acad Sci U S A, 2009. **106**(48): p. 20365-70.

56. Sridhar, S., et al., *Cellular immune correlates of protection against symptomatic pandemic influenza*. Nat Med, 2013. **19**(10): p. 1305-12.

57. Kundu, R., et al., *Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts*. Nat Commun, 2022. **13**(1): p. 80.

58. Ansari, A., et al., *Immune Memory in Mild COVID-19 Patients and Unexposed Donors Reveals Persistent T Cell Responses After SARS-CoV-2 Infection*. Front Immunol, 2021. **12**: p. 636768.

59. Bresciani, A., et al., *T-cell recognition is shaped by epitope sequence conservation in the host proteome and microbiome*. Immunology, 2016. **148**(1): p. 34-9.

60. Klein, L., et al., *Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see)*. Nat Rev Immunol, 2014. **14**(6): p. 377-91.

61. Dong, E., H. Du, and L. Gardner, *An interactive web-based dashboard to track COVID-19 in real time*. Lancet Infect Dis, 2020. **20**(5): p. 533-534.

62. Cutler, D.M. and L.H. Summers, *The COVID-19 Pandemic and the $16 Trillion Virus*. JAMA, 2020. **324**(15): p. 1495-1496.

63. Zhang, Z., et al., *Humoral and cellular immune memory to four COVID-19 vaccines*. Cell, 2022. **185**(14): p. 2434-2451 e17.

64. Liu, L., et al., *Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2*. Nature, 2022. **602**(7898): p. 676-681.

65. Wang, P., et al., *Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7*. Nature, 2021. **593**(7857): p. 130-135.

66. Abu-Raddad, L.J., et al., *Effectiveness of the BNT162b2 Covid-19 Vaccine against the B.1.1.7 and B.1.351 Variants*. N Engl J Med, 2021. **385**(2): p. 187-189.

67. Madhi, S.A., et al., *Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant*. N Engl J Med, 2021. **384**(20): p. 1885-1898.

68. Zhuang, Z., et al., *Mapping and role of T cell response in SARS-CoV-2-infected mice*. J Exp Med, 2021. **218**(4).

69. Zhao, J., J. Zhao, and S. Perlman, *T cell responses are required for protection from clinical disease and for virus clearance in severe acute respiratory syndrome coronavirus-infected mice*. J Virol, 2010. **84**(18): p. 9318-25.

70. McMahan, K., et al., *Correlates of protection against SARS-CoV-2 in rhesus macaques*. Nature, 2021. **590**(7847): p. 630-634.

71. Tan, M., et al., *Immunopathological characteristics of coronavirus disease 2019 cases in Guangzhou, China*. Immunology, 2020. **160**(3): p. 261-268.

72. Goga, A., et al., *Breakthrough SARS-CoV-2 infections during periods of delta and omicron predominance, South Africa*. Lancet, 2022. **400**(10348): p. 269-271.

73. WHO. *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020*. 2000 9/21/2023]; Available from: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

74. Grifoni, A., 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): p. 1489-1501 e15.

75. Li, C.K., et al., *T cell responses to whole SARS coronavirus in humans*. J Immunol, 2008. **181**(8): p. 5490-500.

76. Yang, F., et al., *Validation of an IFN-gamma ELISpot assay to measure cellular immune responses against viral antigens in non-human primates*. Gene Ther, 2022. **29**(1-2): p. 41-54.

77. Slota, M., et al., *ELISpot for measuring human immune responses to vaccines*. Expert Rev Vaccines, 2011. **10**(3): p. 299-306.

78. Bui, H.H., et al., *Predicting population coverage of T-cell epitope-based diagnostics and vaccines*. BMC Bioinformatics, 2006. **7**: p. 153.

79. Teng, J.M. and K.T. Hogan, *Both major and minor peptide-binding pockets in HLA-A2 influence the presentation of influenza virus matrix peptide to cytotoxic T lymphocytes*. Mol Immunol, 1994. **31**(6): p. 459-70.

80. Teng, J.M. and K.T. Hogan, *Residues outside of the HLA-A2 peptide-binding groove can abrogate or enhance recognition of influenza virus matrix peptide pulsed cells by cytotoxic T lymphocytes*. Mol Immunol, 1994. **31**(6): p. 445-57.

81. Jin, X., X. Liu, and C. Shen, *A systemic review of T-cell epitopes defined from the proteome of SARS-CoV-2*. Virus Res, 2023. **324**: p. 199024.

82. Snyder, T.M., et al., *Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels*. medRxiv, 2020.

83. Gonzalez-Galarza, F.F., et al., *Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools*. Nucleic Acids Res, 2020. **48**(D1): p. D783-D788.

84. Mateus, J., et al., *Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans*. Science, 2020. **370**(6512): p. 89-94.

85. Dan, J.M., et al., *Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection*. Science, 2021. **371**(6529).

86. Becker, J.P. and A.B. Riemer, *The Importance of Being Presented: Target Validation by Immunopeptidomics for Epitope-Specific Immunotherapies*. Front Immunol, 2022. **13**: p. 883989.

87. Ferretti, A.P., 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): p. 1095-1107 e3.

88. Weingarten-Gabbay, S., et al., *Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs*. Cell, 2021. **184**(15): p. 3962-3980 e17.

89. Abd El-Baky, N., A.A. Amara, and E.M. Redwan, *HLA-I and HLA-II Peptidomes of SARS-CoV-2: A Review*. Vaccines (Basel), 2023. **11**(3).

90. Tarke, A., et al., *Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases*. bioRxiv, 2020.

91. Abbasian, M.H., et al., *Global landscape of SARS-CoV-2 mutations and conserved regions*. J Transl Med, 2023. **21**(1): p. 152.

92. Saldivar-Espinoza, B., et al., *The Mutational Landscape of SARS-CoV-2*. Int J Mol Sci, 2023. **24**(10).

93. Sette, A. and J. Sidney, *HLA supertypes and supermotifs: a functional perspective on HLA polymorphism*. Curr Opin Immunol, 1998. **10**(4): p. 478-82.

94. Naranbhai, V., et al., *T cell reactivity to the SARS-CoV-2 Omicron variant is preserved in most but not all prior infected and vaccinated individuals*. medRxiv, 2022.

95. Tarke, A., et al., *SARS-CoV-2 vaccination induces immunological T cell memory able to cross-recognize variants from Alpha to Omicron*. Cell, 2022. **185**(5): p. 847-859 e11.

96. Zhang, Y., et al., *Safety, tolerability, and immunogenicity of an inactivated SARS-CoV-2 vaccine in healthy adults aged 18-59 years: a randomised, double-blind, placebo-controlled, phase 1/2 clinical trial*. Lancet Infect Dis, 2021. **21**(2): p. 181-192.

97. Zhu, F.C., et al., *Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial*. Lancet, 2020. **395**(10240): p. 1845-1854.

98. Liu, A.Y., et al., *High background in ELISpot assays is associated with elevated levels of immune activation in HIV-1-seronegative individuals in Nairobi*. Immun Inflamm Dis, 2018. **6**(3): p. 392-401.

99. Ramachandran, N., et al., *Self-assembling protein microarrays*. Science, 2004. **305**(5680): p. 86-90.

100. Madeira, F., et al., *Search and sequence analysis tools services from EMBL-EBI in 2022*. Nucleic Acids Res, 2022. **50**(W1): p. W276-W279.

101. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench*. Bioinformatics, 2009. **25**(9): p. 1189-91.

102. Krishna, S., et al., *Human Papilloma Virus Specific Immunogenicity and Dysfunction of CD8(+) T Cells in Head and Neck Cancer*. Cancer Res, 2018. **78**(21): p. 6159-6170.

103. Havenar-Daughton, C., et al., *Cytokine-Independent Detection of Antigen-Specific Germinal Center T Follicular Helper Cells in Immunized Nonhuman Primates Using a Live Cell Activation-Induced Marker Technique*. J Immunol, 2016. **197**(3): p. 994-1002.

104. Nichols, W.G., A.J. Peck Campbell, and M. Boeckh, *Respiratory viruses other than influenza virus: impact and therapeutic advances*. Clin Microbiol Rev, 2008. **21**(2): p. 274-90, table of contents.

105. Ursin, R.L. and S.L. Klein, *Sex Differences in Respiratory Viral Pathogenesis and Treatments*. Annu Rev Virol, 2021. **8**(1): p. 393-414.

106. Adams, J.J., et al., *Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity*. Nat Immunol, 2016. **17**(1): p. 87-94.

107. Wen, J., et al., *CD4(+) T Cells Cross-Reactive with Dengue and Zika Viruses Protect against Zika Virus Infection*. Cell Rep, 2020. **31**(4): p. 107566.

108. Urbani, S., et al., *Heterologous T cell immunity in severe hepatitis C virus infection*. J Exp Med, 2005. **201**(5): p. 675-80.

109. Cui, J., F. Li, and Z.L. Shi, *Origin and evolution of pathogenic coronaviruses*. Nat Rev Microbiol, 2019. **17**(3): p. 181-192.

110. Le Bert, N., et al., *SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls*. Nature, 2020. **584**(7821): p. 457-462.

111. Cao, W.C., et al., *Disappearance of antibodies to SARS-associated coronavirus after recovery*. N Engl J Med, 2007. **357**(11): p. 1162-3.

112. Nickbakhsh, S., et al., *Epidemiology of Seasonal Coronaviruses: Establishing the Context for the Emergence of Coronavirus Disease 2019*. J Infect Dis, 2020. **222**(1): p. 17-25.

113. Weiskopf, D., et al., *HLA-DRB1 Alleles Are Associated With Different Magnitudes of Dengue Virus-Specific CD4+ T-Cell Responses*. J Infect Dis, 2016. **214**(7): p. 1117-24.

114. Titov, A., et al., *Immunogenic epitope panel for accurate detection of non-cross-reactive T cell response to SARS-CoV-2*. JCI Insight, 2022. **7**(9).

115. Oberhardt, V., et al., *Rapid and stable mobilization of CD8(+) T cells by SARS-CoV-2 mRNA vaccine*. Nature, 2021. **597**(7875): p. 268-273.

116. Francis, J.M., et al., *Allelic variation in class I HLA determines CD8(+) T cell repertoire shape and cross-reactive memory responses to SARS-CoV-2*. Sci Immunol, 2022. **7**(67): p. eabk3070.

117. Somogyi, E., et al., *A Peptide Vaccine Candidate Tailored to Individuals' Genetics Mimics the Multi-Targeted T Cell Immunity of COVID-19 Convalescent Subjects*. Front Genet, 2021. **12**: p. 684152.

118. Prakash, S., et al., *Genome-Wide B Cell, CD4(+), and CD8(+) T Cell Epitopes That Are Highly Conserved between Human and Animal Coronaviruses, Identified from SARS-CoV-2 as Targets for Preemptive Pan-Coronavirus Vaccines*. J Immunol, 2021. **206**(11): p. 2566-2582.

119. Jin, X., et al., *Screening HLA-A-restricted T cell epitopes of SARS-CoV-2 and the induction of CD8(+) T cell responses in HLA-A transgenic mice*. Cell Mol Immunol, 2021. **18**(12): p. 2588-2608.

120. Sievers, F., et al., *Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega*. Mol Syst Biol, 2011. **7**: p. 539.

121. Wang, N., et al., *Structure of MERS-CoV spike receptor-binding domain complexed with human receptor DPP4*. Cell Res, 2013. **23**(8): p. 986-93.

122. Le Bert, N.A.T.T., Kamini Kunasegaran, Christine Y L Tham,, et al., *different pattern of pre-existing SARS-CoV-2 T cell immunity in SARS-recovered and uninfected individuals*. BioRxiv, 2020.

123. Braun, J., et al., *SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19*. Nature, 2020. **587**(7833): p. 270-274.

124. Weiskopf, D., et al., *Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome*. Sci Immunol, 2020. **5**(48).

125. Oseroff, C., et al., *Molecular determinants of T cell epitope recognition to the common Timothy grass allergen*. J Immunol, 2010. **185**(2): p. 943-55.

126. Blum, J.S., P.A. Wearsch, and P. Cresswell, *Pathways of antigen processing*. Annu Rev Immunol, 2013. **31**: p. 443-73.

127. Wolfl, M. and P.D. Greenberg, *Antigen-specific activation and cytokine-facilitated expansion of naive, human CD8+ T cells*. Nat Protoc, 2014. **9**(4): p. 950-66.

128. Gorse, G.J., et al., *Prevalence of antibodies to four human coronaviruses is lower in nasal secretions than in serum*. Clin Vaccine Immunol, 2010. **17**(12): p. 1875-80.

129. Tanaka, T., et al., *Transfer of in vitro-expanded naive T cells after lymphodepletion enhances antitumor immunity through the induction of polyclonal antitumor effector T cells*. PLoS One, 2017. **12**(8): p. e0183976.

130. Rudolph, M., Stanfield Robyn, Wilson Ian, *How TCRs bind MHCs, peptides and coreceptors*. Annual Review of Immunology, 2006. **24**: p. 419-466.

131. Sewell, A.K., *Why must T cells be cross-reactive?* Nat Rev Immunol, 2012. **12**(9): p. 669-77.

132. Petrova, G., A. Ferrante, and J. Gorski, *Cross-reactivity of T cells and its role in the immune system*. Crit Rev Immunol, 2012. **32**(4): p. 349-72.

133. Arstila, T.P., et al., *A direct estimate of the human alphabeta T cell receptor diversity*. Science, 1999. **286**(5441): p. 958-61.

134. Wooldridge, L., et al., *A single autoimmune T cell receptor recognizes more than a million different peptides*. J Biol Chem, 2012. **287**(2): p. 1168-77.

135. Laydon, D.J., C.R. Bangham, and B. Asquith, *Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach*. Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1675).

136. Huang, H., et al., *Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening*. Nat Biotechnol, 2020. **38**(10): p. 1194-1202.

137. Zhang, Y., et al., *Transduction of human T cells with a novel T-cell receptor confers anti-HCV reactivity*. PLoS Pathog, 2010. **6**(7): p. e1001018.

138. Rasmussen, M., et al., *Uncovering the peptide-binding specificities of HLA-C: a general strategy to determine the specificity of any MHC class I molecule*. J Immunol, 2014. **193**(10): p. 4790-802.

139. Thomsen, M.C. and M. Nielsen, *Seq2Logo: a method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion*. Nucleic Acids Res, 2012. **40**(Web Server issue): p. W281-7.

140. Sabatino, J.J., Jr., et al., *High prevalence of low affinity peptide-MHC II tetramer-negative effectors during polyclonal CD4+ T cell responses*. J Exp Med, 2011. **208**(1): p. 81-90.

141. Bacher, P., et al., *Low-Avidity CD4(+) T Cell Responses to SARS-CoV-2 in Unexposed Individuals and Humans with Severe COVID-19*. Immunity, 2020. **53**(6): p. 1258-1271 e5.

142. Dykema, A.G., et al., *Functional characterization of CD4+ T cell receptors crossreactive for SARS-CoV-2 and endemic coronaviruses*. J Clin Invest, 2021. **131**(10).

143. Zehn, D., S.Y. Lee, and M.J. Bevan, *Complete but curtailed T-cell response to very low-affinity antigen*. Nature, 2009. **458**(7235): p. 211-4.

144. Spear, T.T., et al., *Understanding TCR affinity, antigen specificity, and cross-reactivity to improve TCR gene-modified T cells for cancer immunotherapy*. Cancer Immunol Immunother, 2019. **68**(11): p. 1881-1889.

145. Ye, J., et al., *IgBLAST: an immunoglobulin variable domain sequence analysis tool*. Nucleic Acids Res, 2013. **41**(Web Server issue): p. W34-40.

146. Hughes, M.S., et al., *Transfer of a TCR gene derived from a patient with a marked antitumor response conveys highly active T-cell effector functions*. Hum Gene Ther, 2005. **16**(4): p. 457-72.

147. Draper, L.M., et al., *Targeting of HPV-16+ Epithelial Cancer Cells by TCR Gene Engineered T Cells Directed against E6*. Clin Cancer Res, 2015. **21**(19): p. 4431-9.

148. Guan, B.W., TL, Shih IeM, *ARID 1A, a factor that promotes formation of SWI/SNF-mediated chromatin remodeling, is a tumor suppressor in gynecologic cancers*. Cancer Research, 2011. **71**(21).

149. Gerald, J., *Weekly Arizona COVID-19 Data Report*. 2022, University of Arizona, Mel and Enid Zuckerman College of Public Health.

150. Gao, Y., et al., *Ancestral SARS-CoV-2-specific T cells cross-recognize the Omicron variant*. Nat Med, 2022. **28**(3): p. 472-476.

151. Sant, A.J., et al., *CD4 T cells in protection from influenza virus: Viral antigen specificity and functional potential*. Immunol Rev, 2018. **284**(1): p. 91-105.

152. Grant, E.J., et al., *Human influenza viruses and CD8(+) T cell responses*. Curr Opin Virol, 2016. **16**: p. 132-142.

153. Rivino, L., et al., *Differential targeting of viral components by CD4+ versus CD8+ T lymphocytes in dengue virus infection*. J Virol, 2013. **87**(5): p. 2693-706.

APPENDIX A

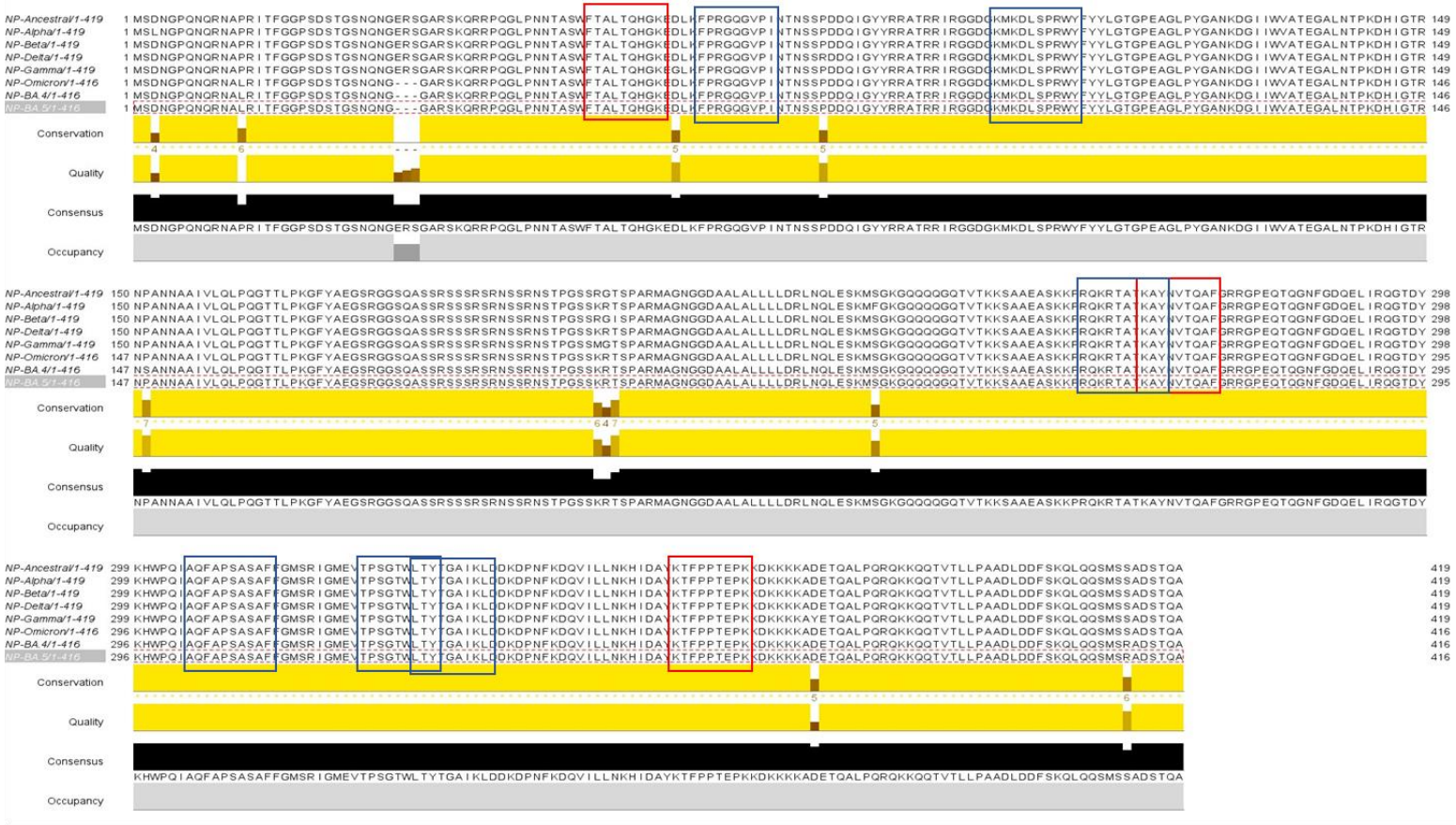
PREDICTED HLA RESTRICTION OF PREVIOUSLY VALIDATED SARS-COV-2 STRUCTURAL EPITOPES

Putative HLA restriction of 6 $CD8^+$ T cell epitopes according to the HLA restriction of the reactive donors as tested by IFN-γ ELISpot secretion assay.

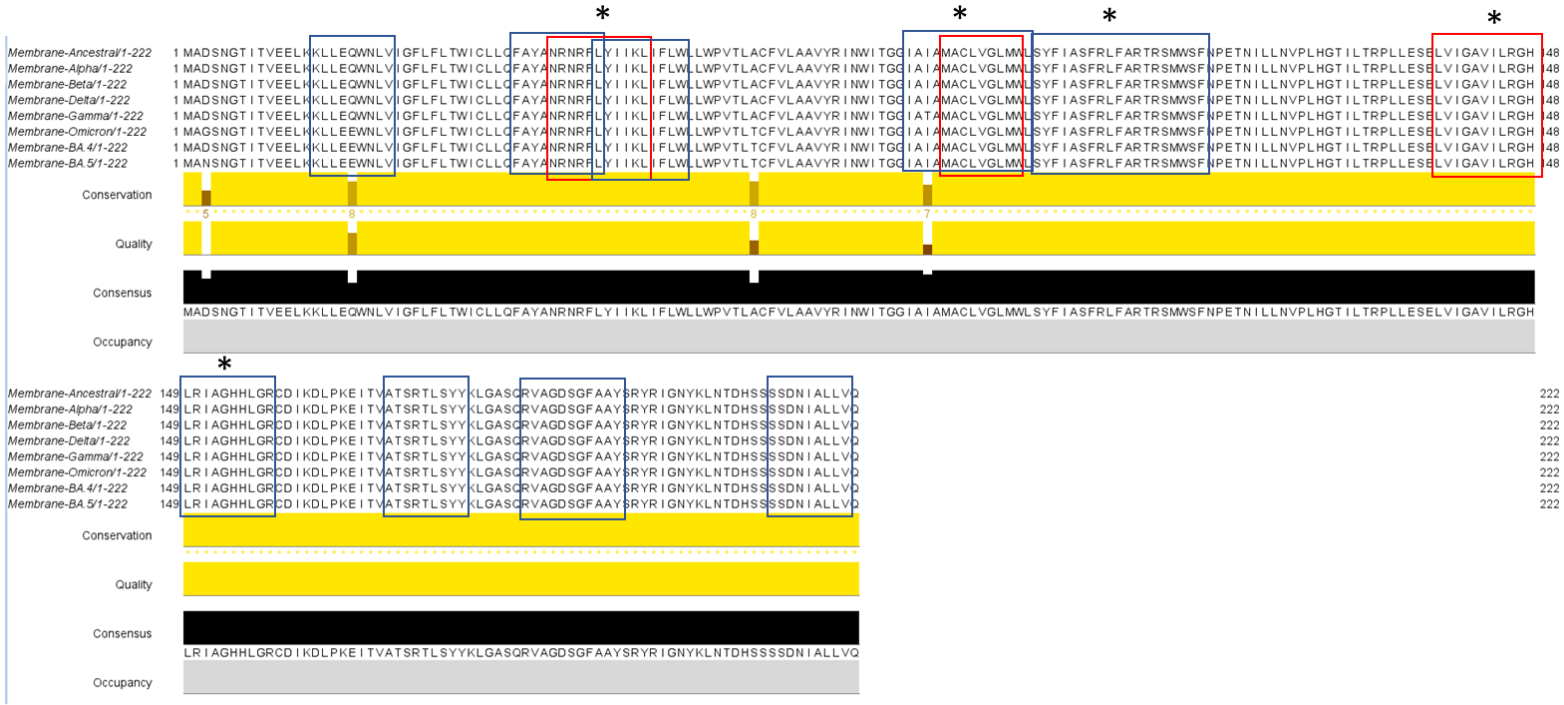
Protein	Peptide Sequence	Start	End	Possible HLA Restriction
E	SLVKPSFYV	70	58	A*02:11
M	EELKKLLEQW	11	21	B*44:02
M	RLFARTRSMWSF	101	112	A*32:01/C*07:01 /C*07:02
S	VYSTGSNVF	630	638	A*24:02
S	IPIGAGICA	659	667	B*35:01/C*05:01/ C*07:01
S	LPFFSNVTWFHA	53	64	C*15:02

APPENDIX B

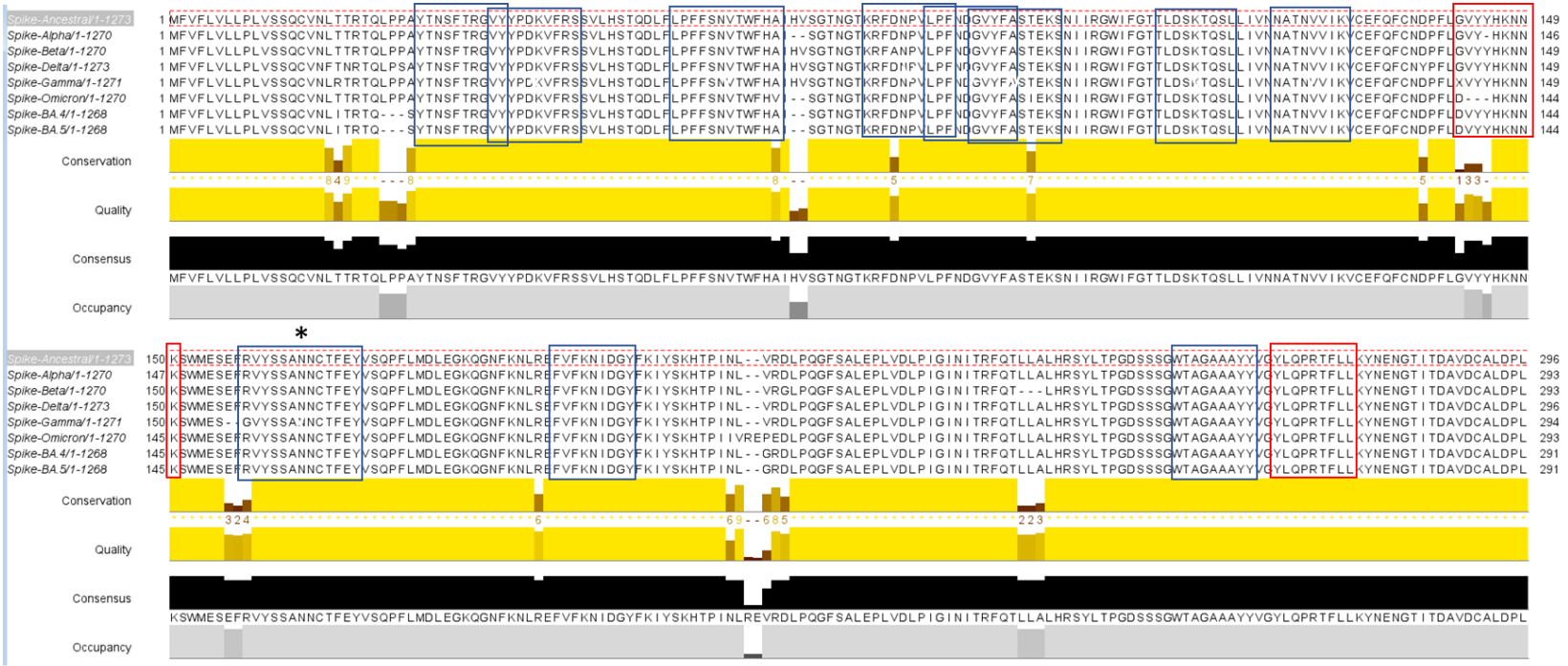
SEQUENCE ALIGNMENT OF VOC SARS-COV-2 STRUCTURAL PROTEINS



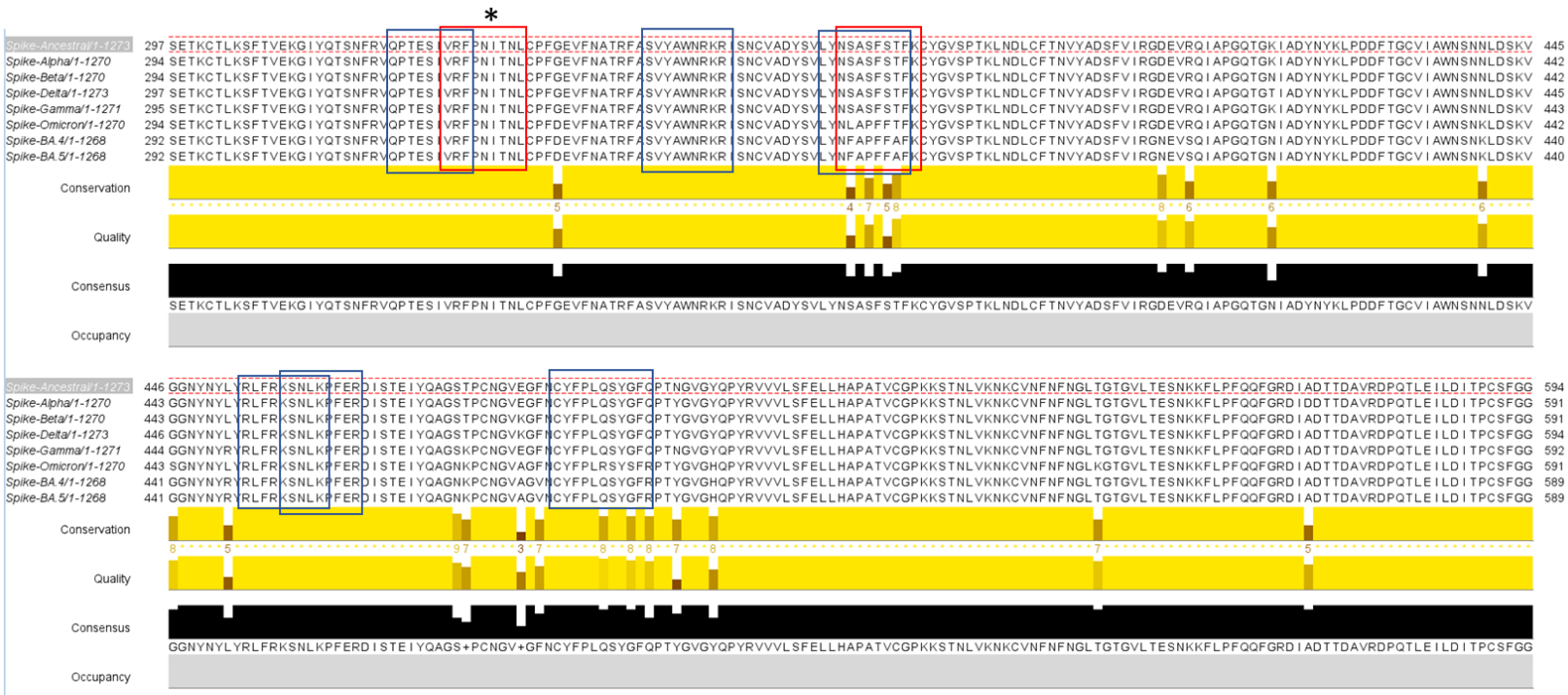
Nucleocapsid amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. * Highlight areas with more than 2 immunogenic and conserved epitopes.



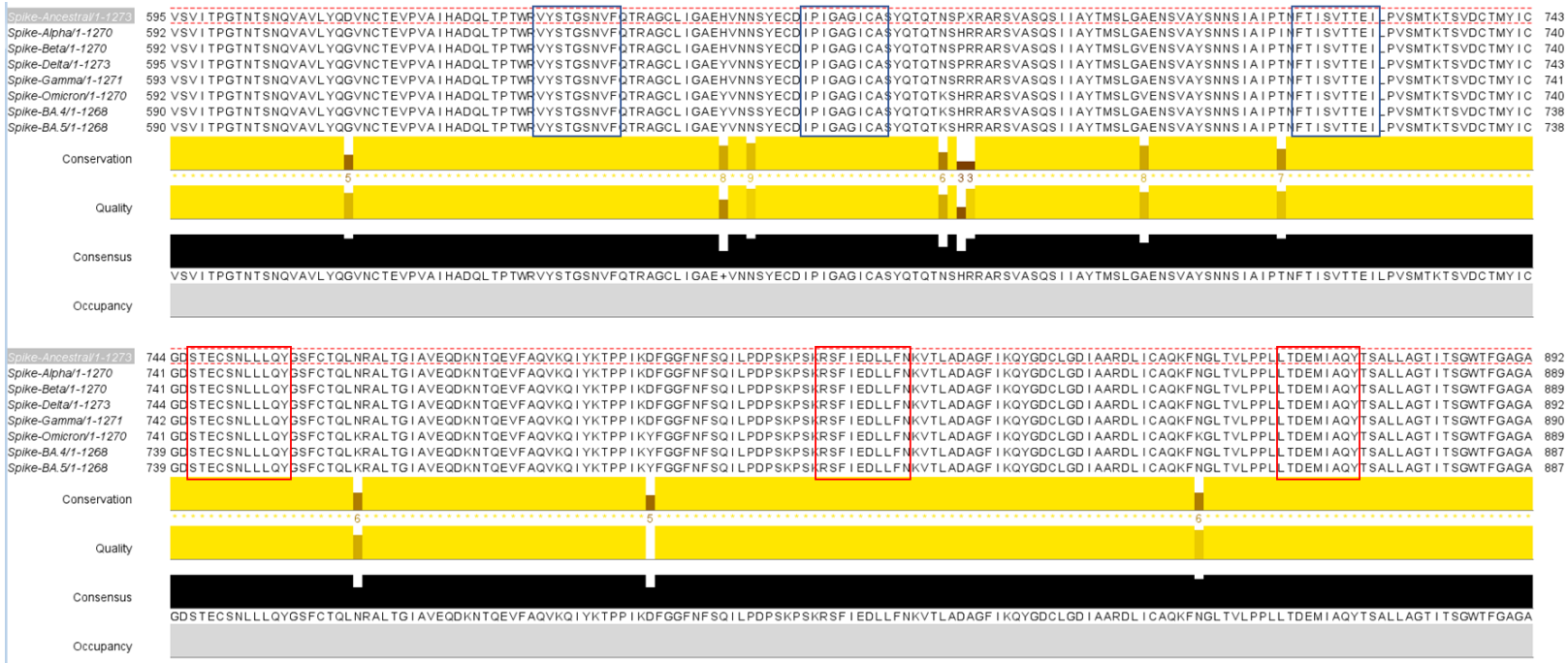
Membrane amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. * Highlight areas with more than 2 immunogenic ad conserved epitopes.



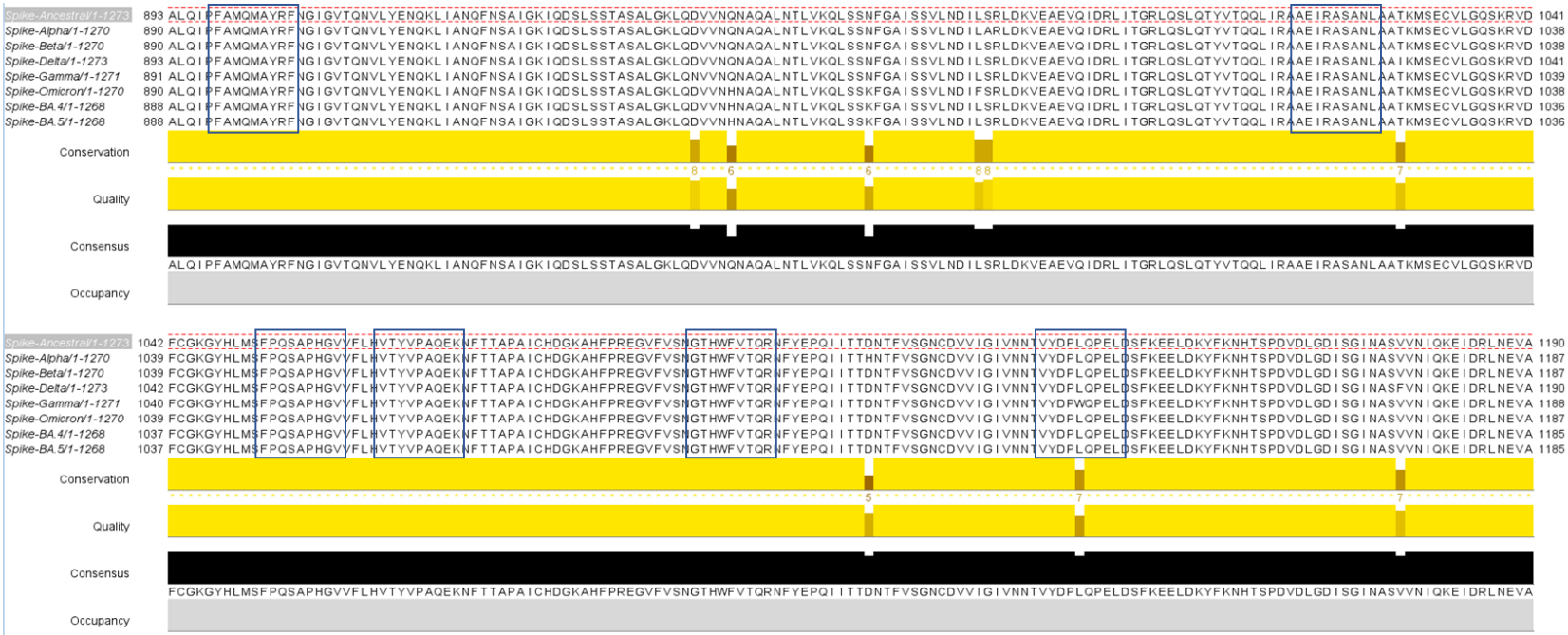
Section 1. Spike amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. * Highlight areas with more than 2 immunogenic ad conserved epitopes. Sequence is divided into 4 sections. Alignments made with Clustal Omega and visualized with JalView.



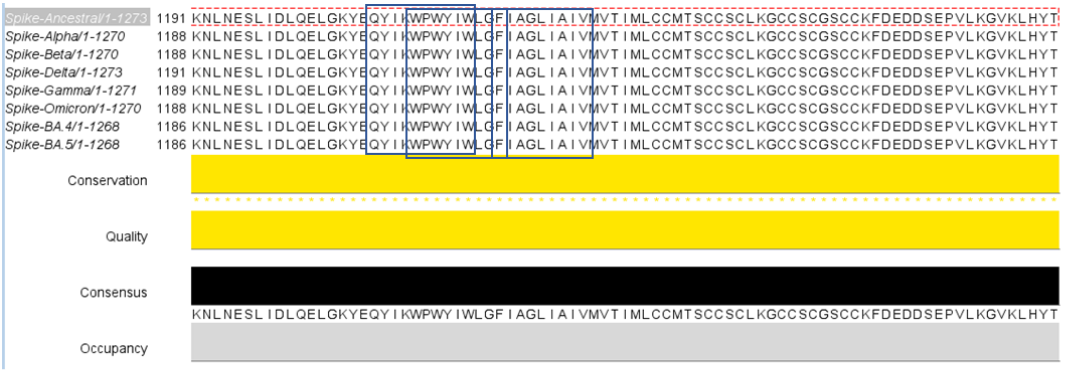
Section 2. Spike amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. * Highlight areas with more than 2 immunogenic ad conserved epitopes. Sequence is divided into 4 sections.



Section 3. Spike amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. * Highlight areas with more than 2 immunogenic ad conserved epitopes. Sequence is divided into 4 sections.



Section 4. Spike amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. (*) Highlight areas with more than 2 immunogenic ad conserved epitopes. Sequence is divided into 4 sections.

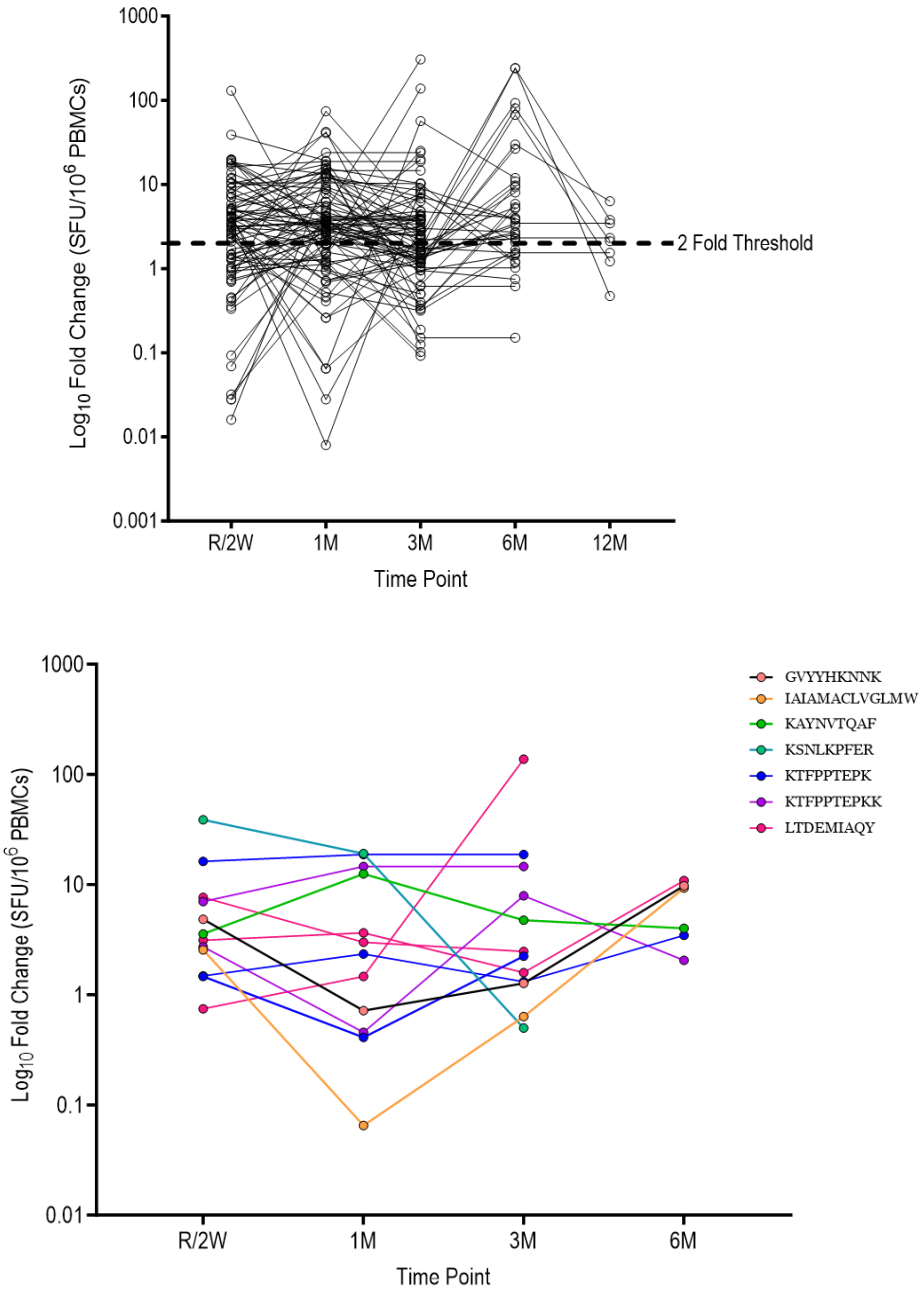


Section 5. Spike amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. * Highlight areas with more than 2 immunogenic ad conserved epitopes. Sequence is divided into 4 sections.

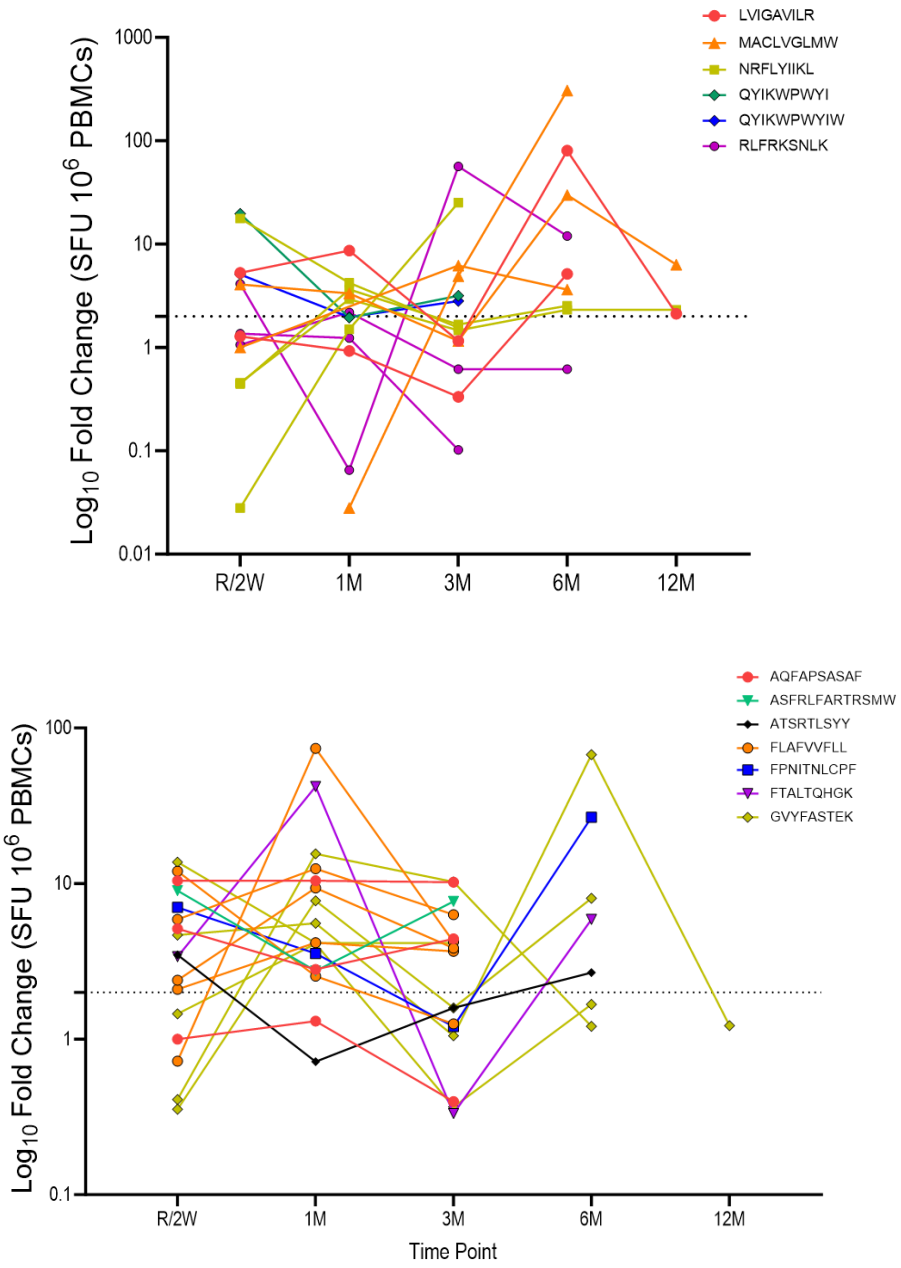
Envelope amino acid alignment of VOC. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Red squares demarcate immunodominant epitopes, blue squares highlight location of other immunogenic epitopes. See Methods for source of protein sequence. * Highlight areas with more than 2 immunogenic and conserved epitopes.

APPENDIX C

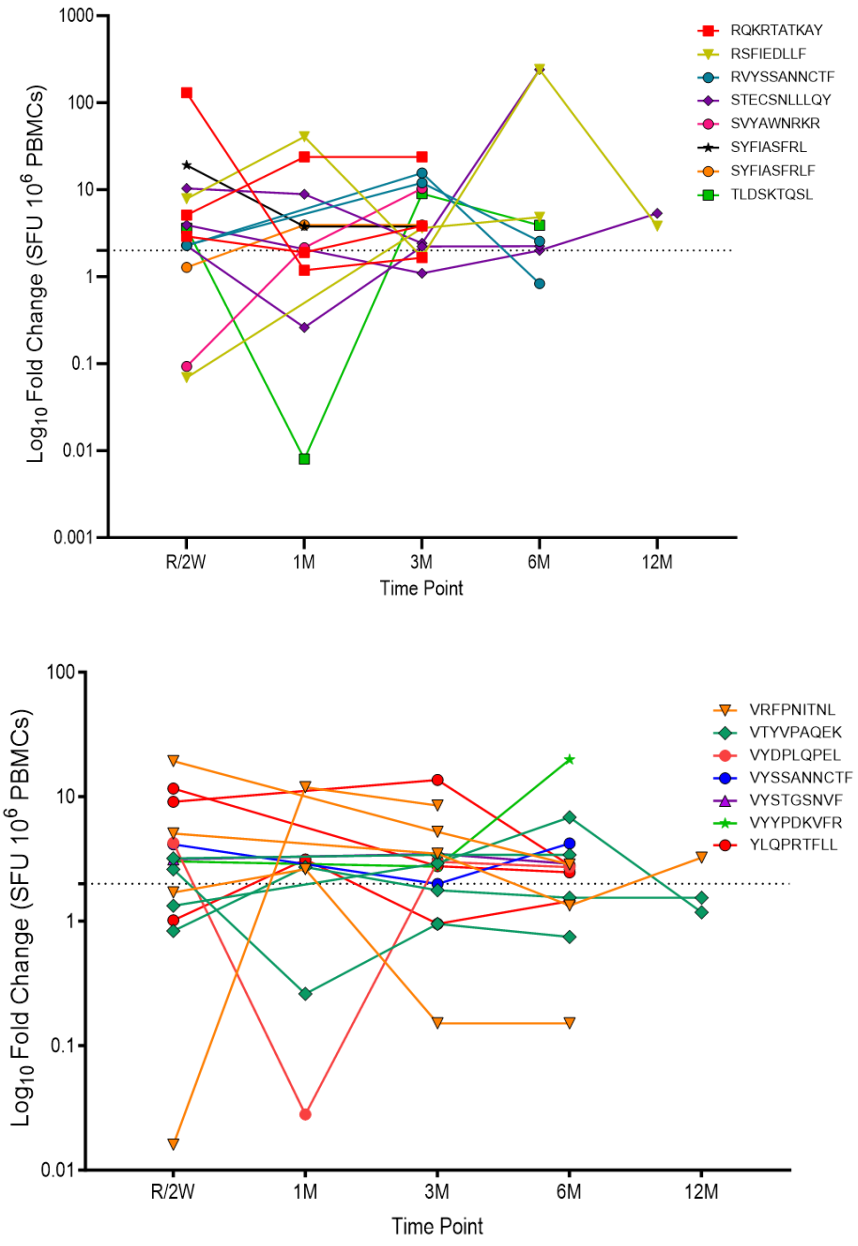
LONGITUDINAL ASSESSMENT OF $CD8^+$ T CELL EPITOPE REACTIVITY



CD8⁺ T cell peptide reactivity over time. Only peptides that were tested over three different time points are plotted. R/2W: Recruitment and 2-week time points were plotted together because most individuals did not have an R time point since their first blood donation was done several weeks/months post-vaccination or infection. 1M/3M/6M/12M: 1 month, 3 months, 6 months and 12 months post-infection or vaccination respectively.



CD8⁺ T cell peptide reactivity over time. Only peptides that were tested over three different time points are plotted. R/2W: Recruitment and 2-week time points were plotted together because most individuals did not have an R time point since their first blood donation was done several weeks/months post-vaccination or infection. 1M/3M/6M/12M: 1 month, 3 months, 6 months, and 12 months post-infection or vaccination respectively.



CD8⁺ T cell peptide reactivity over time. Only peptides that were tested over three different time points are plotted. R/2W: Recruitment and 2-week time points were plotted together because most individuals did not have an R time point since their first blood donation was done several weeks/months post-vaccination or infection. 1M/3M/6M/12M: 1 month, 3 months, 6 months, and 12 months post infection or vaccination respectively.

APPENDIX D

IEDB IMMUNOGENICITY TOOL PREDICTIONS-I

Predicted immunogenicity using the IEDB immunogenicity tool. A negative value indicates the epitope is not predicted to be immunogenic. Reactive column: Result from this study.

Peptide Sequence	Length	IEDB Score	Reactive	PROTEIN	Reported
KKFLPFQQF	9	-0.0392	No	S	N
LPFFSNVTWF	10	0.18944	No	S	N
LPNNTASW	8	-0.0736	No	N	N
LPYGANKDGIIW	12	0.09828	No	N	N
NTQEVFAQV	9	0.17889	No	S	N
SEFRVYSSA	9	-0.10948	No	S	N
TEVPVAIHA	9	0.21049	No	S	N
VATSRTLSY	9	-0.17295	No	M	N
ASFRLFARTRSMW	13	0.12608	Yes	M	N
CYFPLQSYGF	10	-0.23581	Yes	S	N
EELKKLLEQW	10	-0.43502	Yes	M	N
FPNITNLCPF	10	0.1009	Yes	S	N
FPRGQGVPI	9	-0.00164	Yes	S	N
IAIAMACLVGLMW	13	-0.15349	Yes	M	N
IPIGAGICA	9	0.22812	Yes	S	N
KRFDNPVLPF	10	0.06624	Yes	S	N
LPFFSNVTW	9	0.04613	No	S	N
LPFFSNVTWFHA	12	0.39261	Yes	S	N
LPFNDGVYFA	10	0.19023	Yes	S	N
LTDEMIAQY	9	0.02757	Yes	S	N
LTYTGAIKL	9	0.09401	Yes	N	N
QYIKWPWYIW	10	0.31425	Yes	S	N
QYIKWPWYIWLGF	13	0.59175	No	S	N
RLFARTRSMWSF	12	-0.02015	Yes	M	N
RQKRTATKAY	10	-0.06462	Yes	N	N
RSFIEDLLF	9	0.27446	Yes	S	N
RVYSSANNCTF	11	-0.32588	Yes	S	N
SSDNIALLV	9	0.15128	Yes	M	N
SYFIASFRL	9	0.18333	Yes	M	N
SYFIASFRLF	10	0.19632	Yes	M	N
YFIASFRLF	9	0.06887	Yes	M	N
YTNSFTRGV	9	0.04545	Yes	S	N
AEIRASANL	9	0.00689	Yes	S	Y
AEIRASANLA	10	-0.00283	No	S	Y
AQFAPSASAF	10	-0.17446	Yes	N	Y
ATSRTLSYYK	10	-0.13563	No	M	Y
AVILRGHLR	9	0.13516	Yes	M	Y
CVADYSVLY	9	-0.09595	Yes	S	Y
ETKCTLKSF	9	-0.37555	No	S	Y
FAMQMAYRF	9	-0.28061	Yes	S	Y
FAYANRNRF	9	0.10537	Yes	M	Y
FIAGLIAIV	9	0.27206	Yes	S	Y
FLAFVVFLL	9	0.30188	Yes	E	Y

Predicted immunogenicity using the IEDB immunogenicity tool. A negative value indicates the epitope is not predicted to be immunogenic. Reactive column: Result from this study.

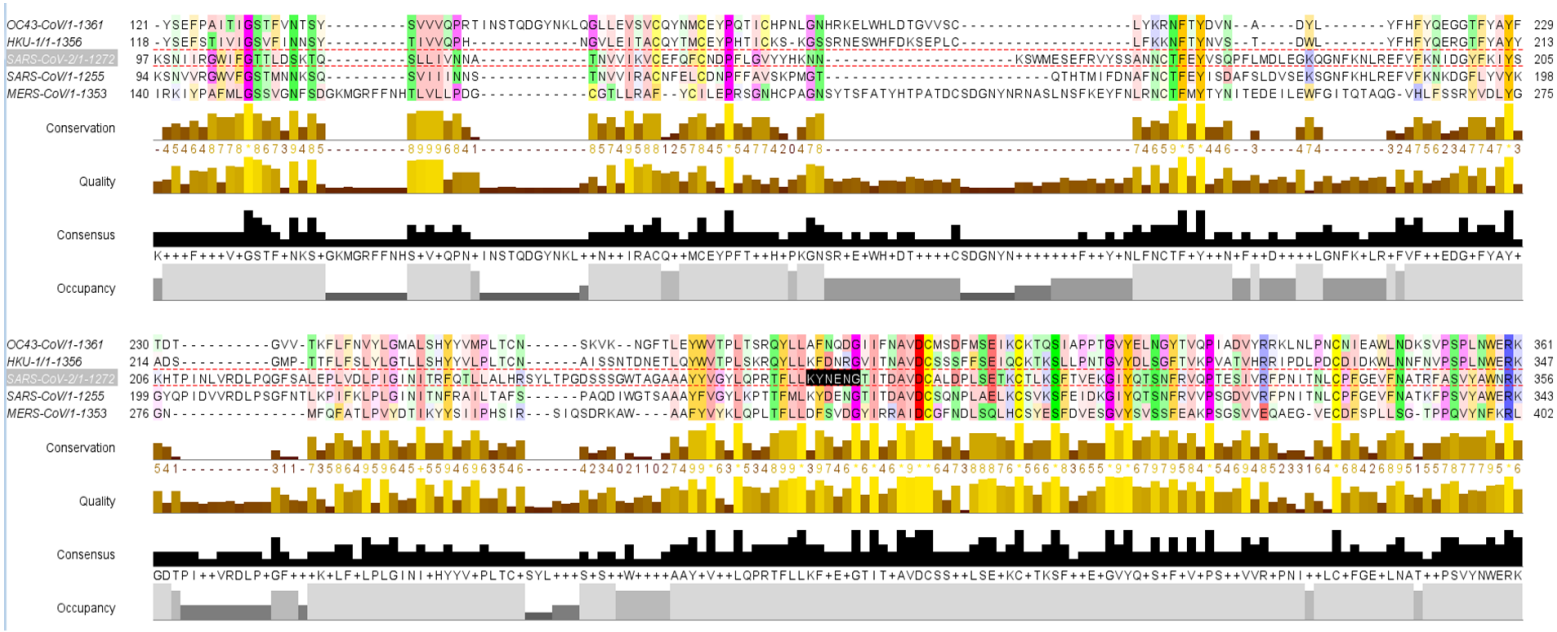
Peptide Sequence	Length	IEDB Score	Reactive	PROTEIN	Reported
FPQSAPHGV	9	-0.12931	Yes	S	Y
FTALTQHGK	9	-0.0226	Yes	N	Y
FTISVTTEI	9	0.04473	Yes	S	Y
FVFKNIDGY	9	-0.0215	Yes	S	Y
GVYFASTEK	9	0.09023	Yes	S	Y
GVYYHKNNK	9	-0.18566	Yes	S	Y
HADQLTPTW	9	-0.0703	No	S	Y
IPFAMQMAY	9	-0.32801	No	S	Y
IPIGAGICASY	11	0.16361	No	S	Y
IPTNFTISV	9	0.17229	No	S	Y
KAYNVTQAF	9	-0.00587	Yes	N	Y
KEIDRLNEV	9	0.15852	No	S	Y
KIADYNYKL	9	-0.10379	No	S	Y
KLLEQWNLV	9	0.18092	Yes	M	Y
KMKDLSPRW	9	-0.19333	Yes	N	Y
KSNLKPFER	9	-0.0764	Yes	S	Y
KTFPPTEPK	9	0.1306	Yes	N	Y
KTFPPTEPKK	10	0.01273	Yes	N	Y
LPFNDGVYF	9	0.11767	Yes	S	Y
LPLVSSQCV	9	-0.40815	No	S	Y
LPNNTASWF	9	0.05582	No	N	Y
LVIGAVILR	9	0.2601	Yes	M	Y
LVKPSFYVY	9	-0.11106	No	E	Y
LYIIKLIFLW	10	0.17392	Yes	M	Y
LYNSASFSTF	10	-0.29831	Yes	S	Y
MACLVGLMW	9	-0.06852	Yes	M	Y
MEVTPSGTW	9	-0.06279	No	N	Y
NATNVVIKV	9	0.07147	Yes	S	Y
NRFLYIIKL	9	0.13484	Yes	M	Y
NSASFSTFK	9	-0.09434	Yes	S	Y
NTASWFTAL	9	0.22775	No	N	Y
QELGKYEQY	9	-0.16616	No	S	Y
QPTESIVRF	9	0.14261	Yes	S	Y
QRNAPRITF	9	0.21019	No	N	Y
QYIKWPWYI	9	0.21624	Yes	S	Y
RIAGHHLGR	9	0.11919	Yes	M	Y
RLFARTRSMW	10	-0.02793	No	M	Y
RLFRKSNLK	9	-0.28759	Yes	S	Y
RVAGDSGFAAY	11	0.10638	Yes	M	Y
RVKNLNSSR	9	-0.32968	Yes	E	Y
SANNCTFEY	9	0.13273	Yes	S	Y
SELVIGAVIL	10	0.34997	Yes	M	Y

Predicted immunogenicity using the IEDB immunogenicity tool. A negative value indicates the epitope is not predicted to be immunogenic. Reactive column: Result from this study.

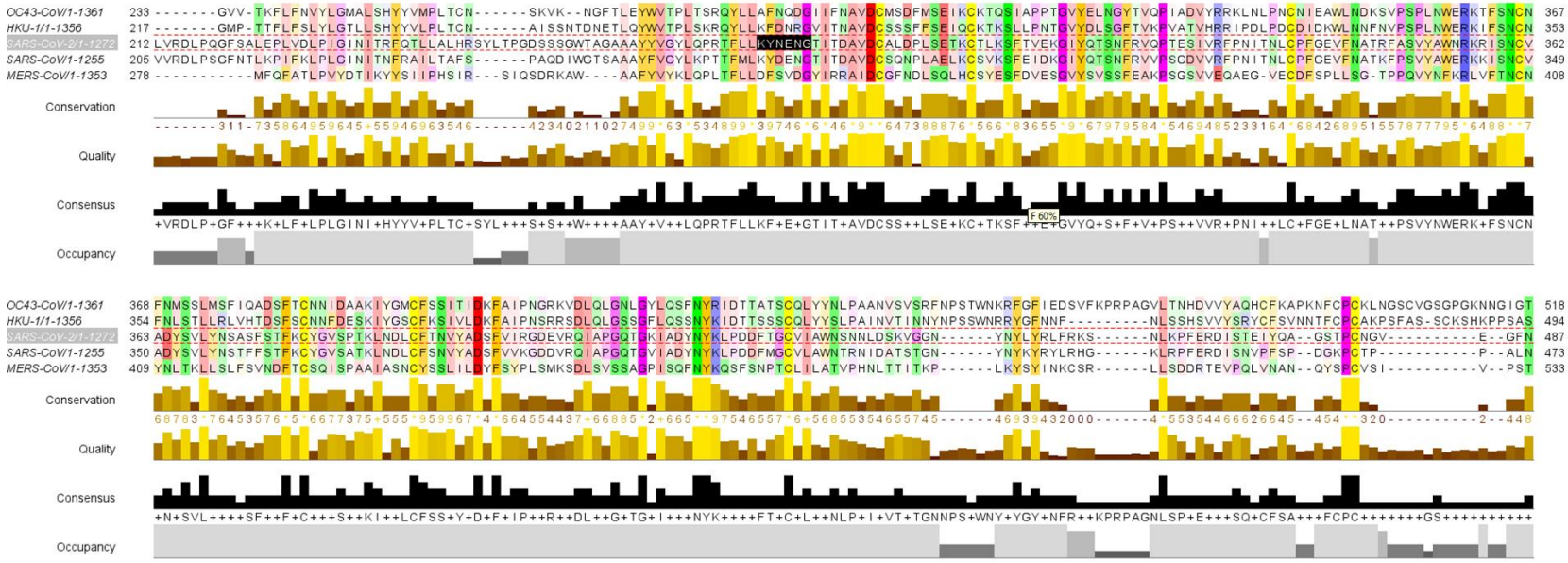
Peptide Sequence	Length	IEDB Score	Reactive	PROTEIN	Reported
SLVKPSFYV	9	-0.27349	Yes	E	Y
STECSNLLLQY	11	-0.28743	Yes	S	Y
SVYAWNRKR	9	0.16546	No	S	Y
TLDSKTQSL	9	-0.52715	Yes	S	Y
TPSGTWLTY	9	0.24003	Yes	N	Y
VASQSIIAY	9	-0.0709	No	S	Y
VLNDILSRL	9	0.03	No	S	Y
VRFPNITNL	9	0.1748	Yes	S	Y
VTYVPAQEK	9	0.02711	Yes	S	Y
VYDPLQPEL	9	-0.07466	Yes	S	Y
VYSSANNCTF	10	-0.21728	Yes	S	Y
VYSTGSNVF	9	-0.11871	Yes	S	Y
VYYPDKVFR	9	-0.09052	Yes	S	Y
WPWYIWLGF	9	0.41673	Yes	S	Y
WTAGAAAYY	9	0.15259	Yes	S	Y
YANRNRFLY	9	0.18472	Yes	M	Y
YLQPRTFLL	9	0.1305	Yes	S	Y
YVYSRVKNL	9	-0.26419	No	E	Y
ATSRTLSYY	9	-0.11604	Yes	M	Y
GTHWFVTQR	9	0.35133	Yes	S	Y
ASANLAATK	9	0.08792	No	S	Y
CPFGEVFNA	9	0.30348	No	S	Y
DISGINASV	9	0.04027	No	S	Y

APPENDIX E

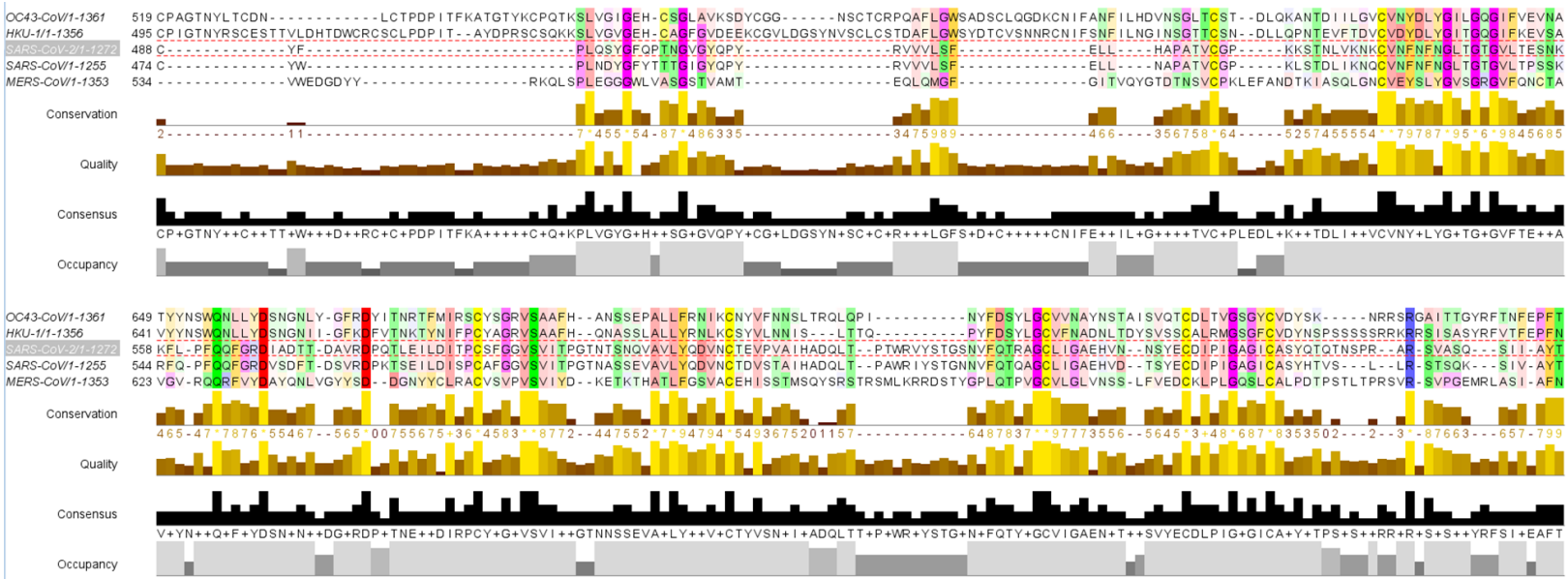
SEQUENCE ALIGNMENT OF SARS-COV-2 AND OTHER HCOVS



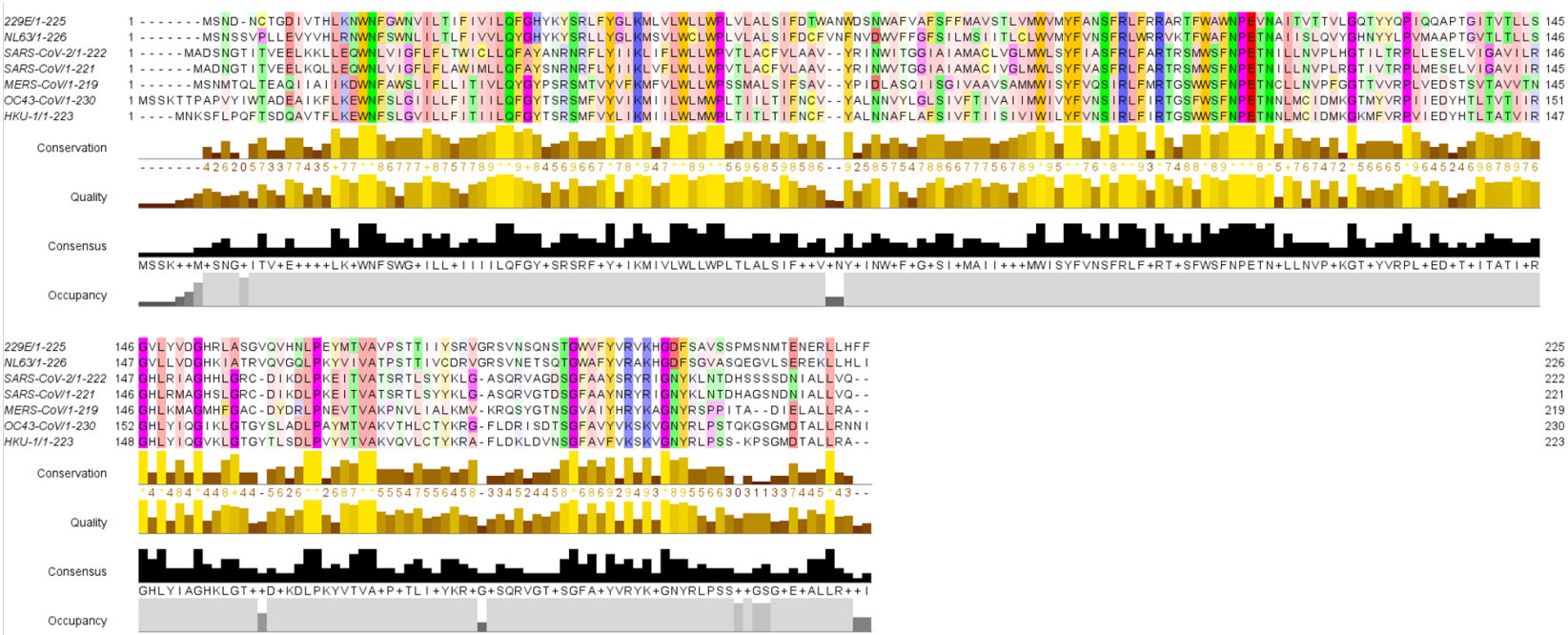
Section 1. Sequence alignment comparing SARS-CoV-2 **spike** protein. Divided into panels. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Zappo and conservation color scheme showing physicochemical property conservation at each amino acid position by intensity of the color. The darker, the more conserved. Pink: Aliphatic/hydrophobic (ILVAM); Orange: Aromatic (FWY); Blue: Positive (KRH); Red: Negative (DE); Green: Hydrophilic (STNQ); Fuscia: conformationally special (PG); Yellow: Cysteine (85). See Methods for source of protein sequence.



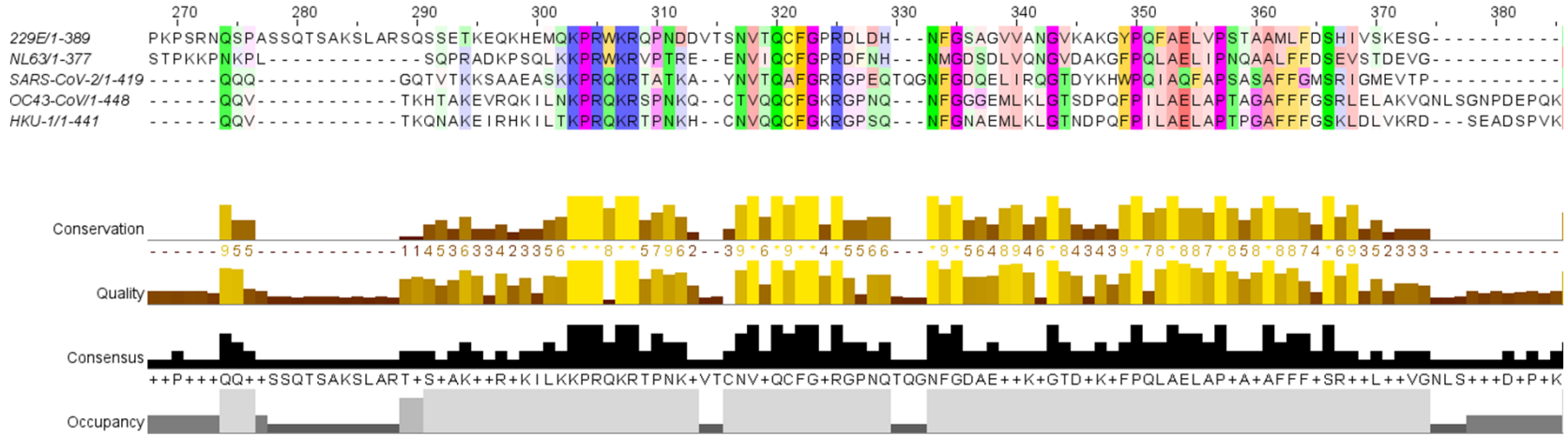
Section 2. Sequence alignment comparing SARS-CoV-2 spike protein. Divided into panels. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Zappo and conservation color scheme showing physicochemical property conservation at each amino acid position by intensity of the color. The darker, the more conserved. Pink: Aliphatic/hydrophobic (ILVAM); Orange: Aromatic (FWY); Blue: Positive (KRH); Red: Negative (DE); Green: Hydrophilic (STNQ); Fuscia: conformationally special (PG); Yellow: Cysteine (85). See Methods for source of protein sequence.



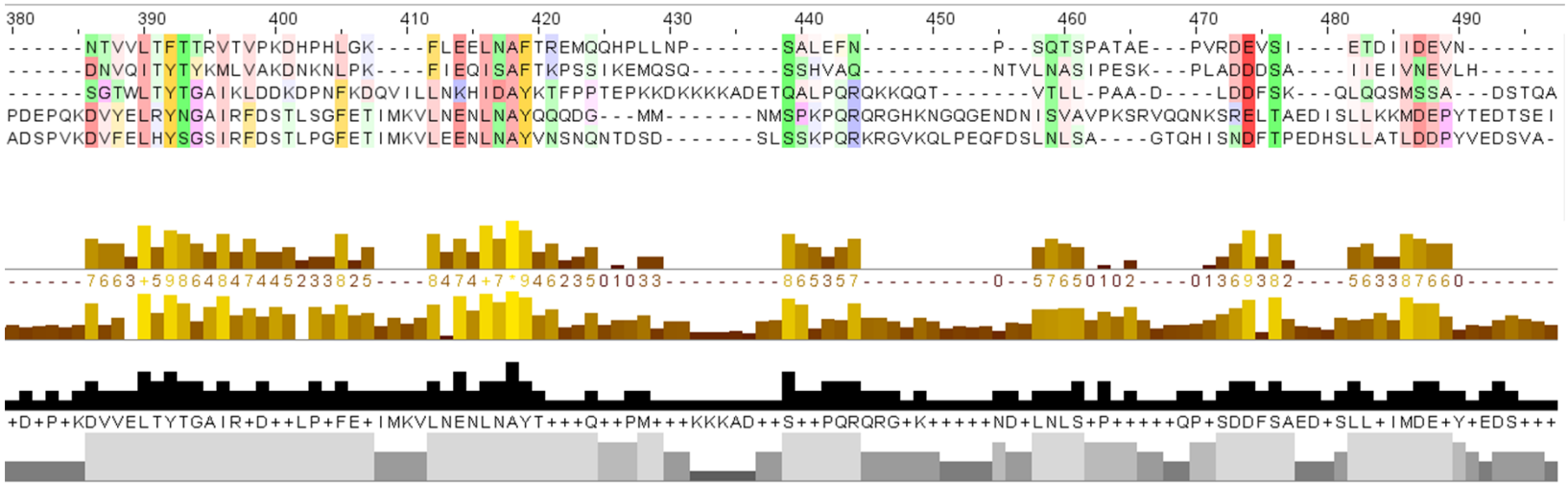
Section 3. Sequence alignment comparing SARS-CoV-2 spike protein. Divided into panels. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Zappo and conservation color scheme showing physicochemical property conservation at each amino acid position by intensity of the color. The darker, the more conserved. Pink: Aliphatic/hydrophobic (ILVAM); Orange: Aromatic (FWY); Blue: Positive (KRH); Red: Negative (DE); Green: Hydrophilic (STNQ); Fuscia: conformationally special (PG); Yellow: Cysteine (85). See Methods for source of protein sequence.



Sequence alignment comparing SARS-CoV-2 Membrane protein. Divided into panels. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Zappo and conservation color scheme showing physicochemical property conservation at each amino acid position by intensity of the color. The darker, the more conserved. Pink: Aliphatic/hydrophobic (ILVAM); Orange: Aromatic (FWY); Blue: Positive (KRH); Red: Negative (DE); Green: Hydrophilic (STNQ); Fuscia: conformationally special (PG); Yellow: Cysteine (85). See Methods for source of protein sequence.



Sequence alignment comparing SARS-CoV-2 Nucleocapsid protein. Divided into panels. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Zappo and conservation color scheme showing physicochemical property conservation at each amino acid position by intensity of the color. The darker, the more conserved. Pink: Aliphatic/hydrophobic (ILVAM); Orange: Aromatic (FWY); Blue: Positive (KRH); Red: Negative (DE); Green: Hydrophilic (STNQ); Fuscia: conformationally special (PG); Yellow: Cysteine (85). See Methods for source of protein sequence.



Sequence alignment comparing SARS-CoV-2 Envelope protein. Divided into panels. Conservation, quality of alignment, consensus residue and deletions are illustrated under the alignment. Zappo and conservation color scheme showing physicochemical property conservation at each amino acid position by intensity of the color. The darker, the more conserved. Pink: Aliphatic/hydrophobic (ILVAM); Orange: Aromatic (FWY); Blue: Positive (KRH); Red: Negative (DE); Green: Hydrophilic (STNQ); Fuscia: conformationally special (PG); Yellow: Cysteine (85). See Methods for source of protein sequence.

APPENDIX F

HUMAN CORONAVIRUS SEQUENCE HOMOLOGY TO SARS-COV-2 IMMUNOGENIC EPITOPES

Part a. Human Coronavirus homologs of SARS-CoV-2 epitopes from S, M, N, and E proteins. Conservation score by amino acid identity and biochemical conservation according to residue properties.

Alpha/Beta	HCoV	Start	End	PROTEIN	Sequence	Conservation Score	Biochemical Conservation	HLA
Alpha	229E				CVVLIVILL			
Alpha	NL-63				LLVMIFFFV			
Beta	SARS-CoV-2	20	28	E	**FLAFVVFLL**	22.3	67%	**A02:11**
Beta	OC-43				VAICLLVTI			
Beta	HKU-1				VLFCLISLI			
Beta	SARS-CoV				FLAFVVFLL			
Beta	MERS-CoV				TVVCAITLL			
Alpha	229E				AVPSTTIIYS			
Alpha	NL-63				ATPSTTIVCD			
Beta	SARS-CoV-2	171	179	M	**ATSRTLSYY**	55.56%	55.56%	**A01:01**
Beta	OC-43				AKVTHLCTYK			
Beta	HKU-1				AKVQVLCTYK			
Beta	SARS-CoV				ATSRTLSYY			
Beta	MERS-CoV				AKPNVLIALK			
Alpha	229E				VTLLSGVLY			
Alpha	NL-63				LTLLSGVLL			
Beta	SARS-CoV-2	142	150	M	**AVILRGHLR**	55.56%	67.00%	**A31:01**
Beta	OC-43				VTIIRGHLY			
Beta	HKU-1				ATVIRGHLY			
Beta	SARS-CoV				AVIIRGHLR			
Beta	MERS-CoV				AVVTNGHLK			
Alpha	229E				-DIVTHLKNW			
Alpha	NL-63				LEVYVHLRNW			
Beta	SARS-CoV-2	11	20	M	**EELKKLLEQW**		67.00%	**B44:02**
Beta	OC-43				DEAIKFLKEW	44.50%		
Beta	HKU-1				DQAVTFLKEW			
Beta	SARS-CoV				EELKQLLEQW			
Beta	MERS-CoV				AQIIAIIKDW			
Alpha	229E				FGHYKYSRL			
Alpha	NL-63				YGHYKYSRL			
Beta	SARS-CoV-2	37	45	M	**FAYANRNRF**		66.70%	**C12:03**
Beta	OC-43				FGYTSRSMF	44.50%		
Beta	HKU-1				FGYTSRSMF			
Beta	SARS-CoV				FAYSNRNRF			
Beta	MERS-CoV				YGYPSRSMT			
Alpha	229E				YVDGHRLASG			
Alpha	NL-65				LVDGHKIATR	33.40%		
Beta	SARS-CoV-4	150	158	M	**RIAGHHLGR**		33.40%	**A31:01**
Beta	OC-45				YIQGIKLGTG			
Beta	HKU-3				YIQGVKLGTG			
Beta	SARS-CoV				RMAGHSLG-R			
Beta	MERS-CoV				KMAGMHFG-A			
Alpha	229E				APTGITVTLL			
Alpha	NL-68				APTGVTLTLL			
Beta	SARS-CoV-7	136	145	M	**SELVIGAVIL**	0% (0/6)		**B40:01**
Beta	OC-48				DYHTLTVTII			
Beta	HKU-6				DYHTLTATVI			
Beta	SARS-CoV				SELVIGAVII			
Beta	MERS-CoV				DSTSVTAVVT			
Alpha	229E				NMTENERLL			
Alpha	NL-69				VLSEREKLL			
Beta	SARS-CoV-8	212	221	M	**SSDNIALLV**		44.50%	**C05:01**
Beta	OC-49				SGMDTALLR	44.50%		
Beta	HKU-7				SGMDTALLR			
Beta	SARS-CoV				**SNDNIALLV**			
Beta	MERS-CoV				-TADIELAL			

Part b. Continued

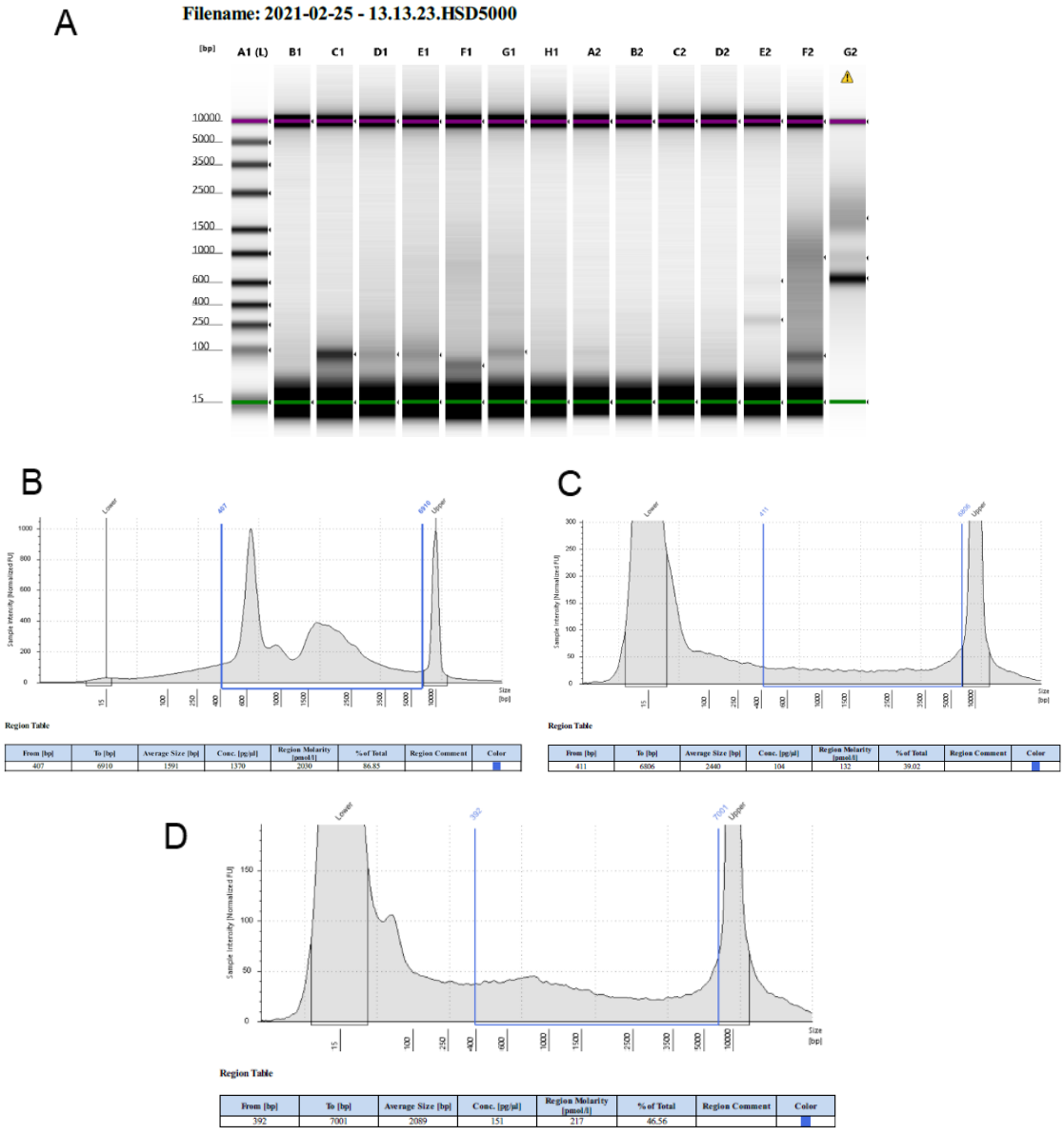
Alpha/Beta	HCoV	Start	End	PROTEIN	Sequence	Conservation Score	Biochemical Conservation	HLA
Alpha	229E				HYKYSRLFY			
Alpha	NL-63				HYKYSRLLY			
Beta	SARS-CoV-2	39	47	M	**YANRNRFLY**		44.50%	A29:02
Beta	OC-43				YTSRSMFVY	44.50%		
Beta	HKU-1				YTSRSMFVY			
Beta	SARS-CoV				YSNRNRFLY			
Beta	MERS-CoV				YPSRSMTVY			
Alpha	229E				DDVTSNVTQCF			
Alpha	NL-63				REENVIQCF			
Beta	SARS-CoV-2	266	274	N	**KAYNVTQAF**		66.70%	B57:01
Beta	OC-43				KQCTVQQCF	55.56%		C12:03
Beta	HKU-1				KHCNVQQCF			
Beta	SARS-CoV				**KQYNVTQAF**			
Beta	MERS-CoV				KSFNMVQAF			
Alpha	229E				TEVRASRQL			
Alpha	NL-63				ESAALSAQL	44.50%		
Beta	SARS-CoV-2	1016	1024	S	**AEIRASANL**		44.50%	B40:01
Beta	OC-43				TLVKFSAAQ			
Beta	HKU-1				TLIKAGASR			
Beta	SARS-CoV				**AEIRASANL**			
Beta	MERS-CoV				ESAALSAQL			
Alpha	229E				-----LY			
Alpha	NL-63				CNYNLTKLL			
Beta	SARS-CoV-2	361	369	S	**CVADYSVLY**		33.40%	A29:02
Beta	OC-43				CNFNMSSLM			
Beta	HKU-1				CNFNLSTLL	33.40%		
Beta	SARS-CoV				CVADYSVLY			
Beta	MERS-CoV				CNYNLTKLL			
Alpha	229E				------			
Alpha	NL-63				------			
Beta	SARS-CoV-2	488	496	S	**CYFPLQSYGF**		22.30%	A23:01
Beta	OC-43				KRFGFIEDSV	16.67% (1/6)		A24:02
Beta	HKU-1				RRYGFGSFNV			
Beta	SARS-CoV				CYWPLNDYGF			
Beta	MERS-CoV				------			
Alpha	229E				------			
Alpha	NL-63				------			
Beta	SARS-CoV-2	458	464	S	**KSNLKPFER**	0% (0/6)	0% (0/6)	A31:01
Beta	OC-43				-RFNP--			
Beta	HKU-1				-NFNP--			
Beta	SARS-CoV				HGKLRPFER			
Beta	MERS-CoV				------			
Alpha	229E				VTFVNISRS			
Alpha	NL-63				NEAFQKVQD			
Beta	SARS-CoV-2	368	376	S	**LYNSASFSTF**	16.67% (1/6)	44.50%	A32:01
Beta	OC-43				ANAFNNALY			A24:02
Beta	HKU-1				QNGFTATNS			A24:01
Beta	SARS-CoV				**LYNS**TF**FSTF**			
Beta	MERS-CoV				NEAFQKBQD			
Alpha	229E				TYIKWPWWV			
Alpha	NL-63				YYNKWPWYI			
Beta	SARS-CoV-2	1208	1216	S	**QYIKWPWYI**		78%	A24:02
Beta	OC-43				YYVKWPWYV	66.67% (4/6)		
Beta	HKU-1				MYVKWPWYI			
Beta	SARS-CoV				QYIKWPWYV			
Beta	MERS-CoV				YYNKWPWYI			

Part c. Continued.

Alpha/Beta	HCoV	Start	End	PROTEIN	Sequence	Conservation Score	Biochemical Conservation	HLA
Alpha	229E				V------			
Alpha	NL-63				------			
Beta	SARS-CoV-2	454	462	S	**RLFRKSNLK**	0% (0/6)	33.40%	**A03:01**
Beta	OC-43				------RFN			
Beta	HKU-1				------NFNP			
Beta	SARS-CoV				RYLRHGKLRP			
Beta	MERS-CoV				---------			
Alpha	229E				VRSFQPLLLNC			
Alpha	NL-63				MGRFFNHTLVL			
Beta	SARS-CoV-2	109	117	S	**TLDSKTQSL**		33.40%	**C05:01**
Beta	OC-43				TQDGYNKLQGL	22.30%		
Beta	HKU-1				---------GIL			
Beta	SARS-CoV				TMN--NKSQSV			
Beta	MERS-CoV				MGRFFNHTLVL			
Alpha	229E				ELQTIVPE-			
Alpha	NL-63				LPPPLLGNS			
Beta	SARS-CoV-2	1137	1145	S	**VYDPLQPEL**	0% (0/6)	33.40%	**C04:01**
Beta	OC-43				MLNTSIPNL			**C05:01**
Beta	HKU-1				YLNNSIPNL			**C14:02**
Beta	SARS-CoV				**VYDPLQPEL**			
Beta	MERS-CoV				LPPPLLGNS			
Alpha	229E				ILFKTSYGVV			
Alpha	NL-63				EYFNLRNCTF			
Beta	SARS-CoV-2	159	167	S	**VYSSANNCTF**	0% (0/6)	33.40%	**A24:02**
Beta	OC-43				LDTGVVSCLY			
Beta	HKU-1				IDSSEPLCLF			
Beta	SARS-CoV				IFDNAFNCTF			
Beta	MERS-CoV				EYFNLRNCTF			
Alpha	229E				ENFTSYGFS			
Alpha	NL-63				RRDSTYGPL	22.30%		
Beta	SARS-CoV-2	635	643	S	**VYSTGSNVF**		0% (0/6)	**A24:02**
Beta	OC-43				------NYF			
Beta	HKU-1				-------YF			
Beta	SARS-CoV				I**YSTGNNVF**			
Beta	MERS-CoV				RRDSTYGPL			
Alpha	229E				ALASYADVL			
Alpha	NL-63				**KLQPLTFLL**			
Beta	SARS-CoV-2	269	277	S	**YLQPRTFLL**	77.80%	88.00%	**A02:11**
Beta	OC-43				PLTSRQYLL			**B08:01**
Beta	HKU-1				PLSRRQYLL			**A02:01**
Beta	SARS-CoV				YLKPTTFML			
Beta	MERS-CoV				K**LQPLTFLL**			

APPENDIX G

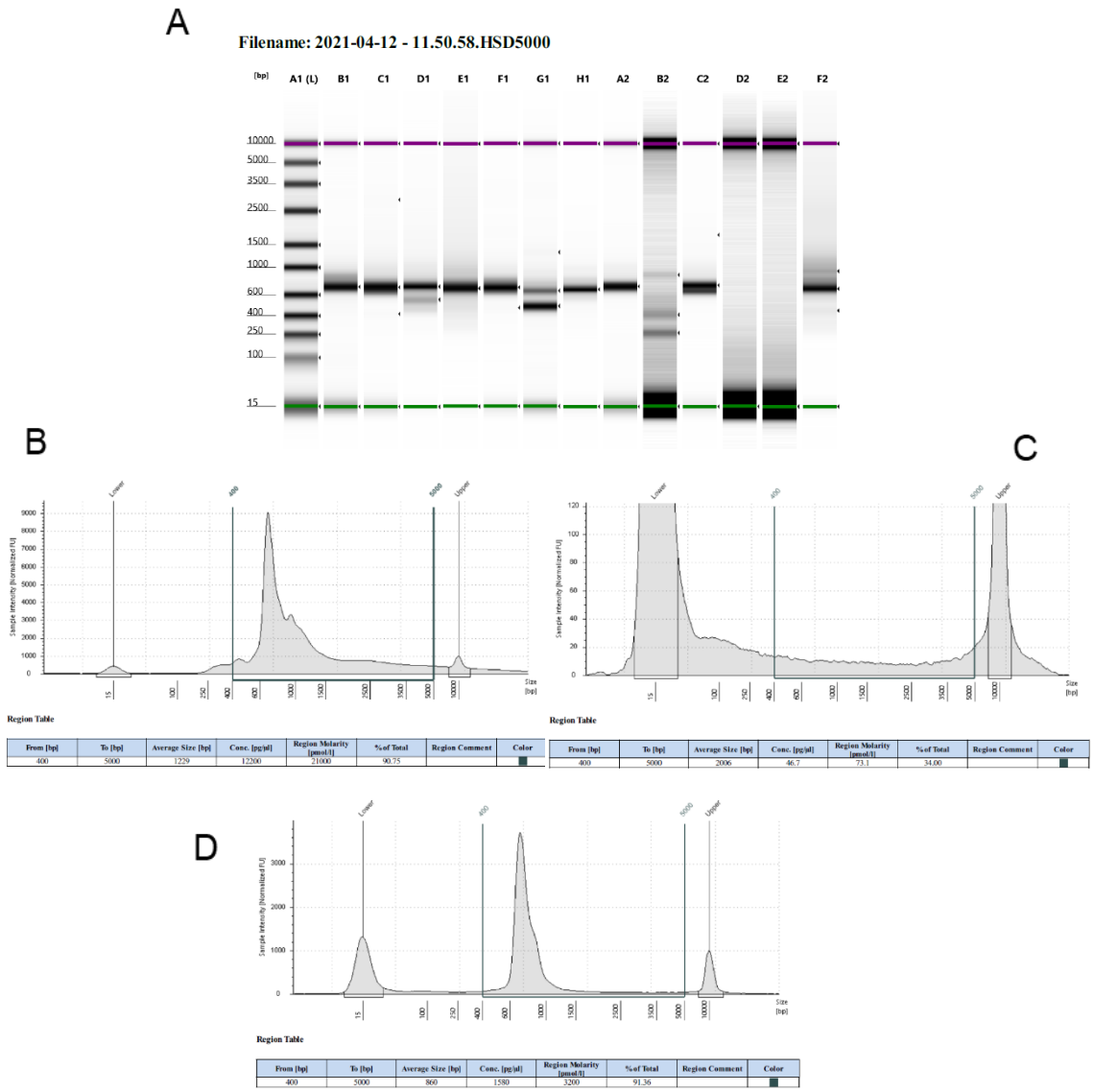
QUALITY CONTROL OF "FIRST STEP" PCR PRODUCTS USING THE SINGLE CELL RNA SEQUENCING TAKARA KIT



(A) Gel electrophoresis of amplified cDNA libraries. Each column is the product of a single pooled column from the 96 well Takara plate. Each well is analyzed with an electropherogram profile to determine the profile (size and abundance) of amplified TCR sequencing libraries. (**B**) Positive control-Jurkat DNA from well G2; (**C**) Negative control; from well F2; (**D**) Sample from well F1, library with peaks from 400 to 3500 bp. TCR α-peak is ~900bp and TCR β at ~700 bp.

APPENDIX H

QUALITY CONTROL OF "SECOND STEP" PCR PRODUCTS USING THE SINGLE CELL RNA SEQUENCING TAKARA KIT



(A) Gel electrophoresis of amplified cDNA libraries with SMART sequence (row barcode) and TCR specific forward and reverse primers (column barcodes). Each well is analyzed with an electropherogram profile to determine the profile (size and abundance) of amplified TCR sequencing libraries, this amplification should amplify the TCR region. (**B**) Positive control-Jurkat DNA from well F2; (**C**) Negative control; from well E2; (**D**) Sample from well B1, library with peaks from 600 to 1000 bp. TCR α-peak is ~900bp and TCR β at ~700 bp.

APPENDIX I

PRODUCTIVE TCR GENE SEQUENCES

TCR αβ pair sequences:

Pair 1:

atgggcaccaggctcctcttctgggtggccttctgtctcctgggggcagatcacacaGGAGCTGGAGTCTCCCAGTC
CCCCAGTAACAAGGTCACAGAGAAGGGAAAGGATGTAGAGCTCAGGTGTGATCCAATT
TCAGGTCATACTGCCCTTTACTGGTACCGACAGAGCCTGGGGCAGGGCCTGGAGTTTT
TGATTTACTTCCAAGGCAACAGTGTACCAGACAAATCAGGGCTGCCCAGTGATCGCTTC
TCTGCAGAGAGGACTGGGGGGTCCGTCTCCACTCTGACGATCCAGCGCACACAGCAGG
AGGACTCGGCCGTGTATCTCTGTGCCAGCAGCTTACTAATTTCCAGGTTTGAAGCTTTC
TTTGGACAAGGCACCAGACTCACAGTTGTAGAGGATCTCCGGAatgtgacacccccctaaggtgtct
ctgttcgagccaagcaaggccgagatcgccaacaagcagaaggccaccctggtgtgcctggccagaggcttctttcccgatcac
gtggagctgtcctggtgggtgaatggcaaggaggtgcactctggcgtgtgcaccgaccctcaggcctataaggagtccaactact
cttattgtctgagctcccggctgagagtgtccgccacattctggcacaatcccagaaaccacttcagatgccaggtgcagtttcacg
gcctgtccgaggaggataagtggcctgagggctctccaaagcccgtgacccagaatatcagcgccgaggcatggggaagggca
gactgtggaatcacctccgcctcttaccagcagggcgtgctgagcgccacaatcctgtatgagatcctgctgggcaaggccaccc
tgtacgccgtgctggtgagcacactggtggtcatggctatggtgaagaggaagaacagccgggccaagcgcagcggctccggc
gcaaccaacttctctctgctgaagcaggcaggcgacgtggaggagaatcctggcccaatgaagaggatattgggagctctgctg
gggctcttgagtgcccaggtttgctgtgtgagaGGAATACAAGTGGAGCAGAGTCCTCCAGACCTGATT
CTCCAGGAGGGAGCCAATTCCACGCTGCGGTGCAATTTTTCTGACTCTGTGAACAATTT
GCAGTGGTTTCATCAAAACCCTTGGGGACAGCTCATCAACCTGTTTTACATTCCCTCAG
GGACAAAACAGAATGGAAGATTAAGCGCCACGACTGTCGCTACGGAACGCTACAGCTT
ATTGTACATTTCCTCTTCCCAGACCACAGACTCAGGCGTTTATTTCTGTGCTGTCGAAC
GGGGCACCGGTAACCAGTTCTATTTTGGGACAGGGACAAGTTTGACGGTCATTCCAAA
TATTCAGAATCCCGAGCCCGCGGtataccagctgaaggacccccggagccaggatagcaccctgtgcctgttcac
agactttgattctcagatcaacgtgcccaagacaatggagagcggcacctttatcacagacaagtgcgtgctggacatgaaggct
atggactctaagagcaatggcgccatcgcctggtccaaccagacctctttcacatgccaggatatctttaaggagacaaatgcca
cataccccagctccgacgtgccttgtgatgccaccctgacagagaagagcttcgagacagacatgaatctgaactttcagaacct
gctggtcatcgtgctgagaatcctgctgctgaaagtggctggcttcaacctgctgatgaccctgcggctgtggagtagctaatga

Pair 2:

atgggcaccaggctcctcttctgggtggccttctgtctcctgggggcagatcacacaGGAGCTGGAGTCTCCCAGTC
CCCCAGTAACAAGGTCACAGAGAAGGGAAAGGATGTAGAGCTCAGGTGTGATCCAATT
TCAGGTCATACTGCCCTTTACTGGTACCGACAGAGCCTGGGGCAGGGCCTGAAGTTTTT
AATTTACTTCCAAGGCAACAGTGCACCAGACAAATCAGGGCTGCCCAGTGATCGCTTCT
CTGCAGAGAGGACTGGGGGATCCGTCTCCACTCTGACGATCCAGCGCACACAGCAGGA
GGACTCGGCCGTGTATCTCTGTGCCAGCAGCTTACTAATTTCCAGGTTTGAAGCTTTCT
TTGGACAAGGCACCAGACTCACAGTTGTAGAGGATCTCCGGAatgtgacacccccctaaggtgtctct
gttcgagccaagcaaggccgagatcgccaacaagcagaaggccaccctggtgtgcctggccagaggcttctttcccgatcacgt
ggagctgtcctggtgggtgaatggcaaggaggtgcactctggcgtgtgcaccgaccctcaggcctataaggagtccaactactct
tattgtctgagctcccggctgagagtgtccgccacattctggcacaatcccagaaaccacttcagatgccaggtgcagtttcacgg
cctgtccgaggaggataagtggcctgagggctctccaaagcccgtgacccagaatatcagcgccgaggcatggggaagggcag
actgtggaatcacctccgcctcttaccagcagggcgtgctgagcgccacaatcctgtatgagatcctgctgggcaaggccaccct
gtacgccgtgctggtgagcacactggtggtcatggctatggtgaagaggaagaacagccgggccaagcgcagcggctccggcg

Pair 2 cntd:
caaccaacttctctctgctgaagcaggcaggcgacgtggaggagaatcctggcccaatggagaccctcttgggcctgcttatcctt
tggctgcagctgcaatgggtgagcagcAAACAGGAGGTGACGCAGATTCCTGCAGCTCTGAGTGTC
CCAGAAGGAGAAAACTTGGTTCTCAACTGCAGTTTCACTGATAGCGCTATTTACAACCT
CCAGTGGTTTAGGCAGGACCCTGGGAAAGGTCTCACATCTCTGTTGCTTATTCAGTCAA
GTCAGAGAGAGCAAACAAGTGGAAGACTTAATGCCTCGCTGGATAAATCATCAGGACG
TAGTACTTTATACATTGCAGCTTCTCAGCCTGGTGACTCAGCCACCTACCTCTGTGCTG
GAGAAACCAGTGGCTCTAGGTTGACCTTTGGGGAAGGAACACAGCTCACAGTGAATCC
TAATATTCAGAATCCCGAGCCCGCGGtataccagctgaaggacccccggagccaggatagcaccctgtgcctg
ttcacagactttgattctcagatcaacgtgcccaagacaatggagagcggcacctttatcacagacaagtgcgtgctggacatga
aggctatggactctaagagcaatggcgccatcgcctggtccaaccagacctctttcacatgccaggatatctttaaggagacaaa
tgccacataccccagctccgacgtgccttgtgatgccaccctgacagagaagagcttcgagacagacatgaatctgaactttcag
aacctgctggtcatcgtgctgagaatcctgctgctgaaagtggctggcttcaacctgctgatgaccctgcggctgtggagtagcta
a

<u>Pair 3</u>:
atgggcccccagctccttggctatgtggtcctttgccttctaggagcaggcccccctgGAAGCCCAAGTGACCCAGAA
CCCAAGATACCTCATCACAGTGACTGGAAAGAAGTTAACAGTGACTTGTTCTCAGAATA
TGAACCATGAGTATATGTTCTGGTATCGACAAGACCCAGGGCTGGGCTTAAGGCAGAT
CTACTATTCAATGAATGTTGAGGTGACTGATAAGGGAGATGTTCCTGAAGGGTACAAA
GTCTCTCGAAAAGAGAAGAGGAATTTCCCCCTGATCCTGGAGTCGCCCAGCCCCAACCA
GACCTCTCTGTACTTCTGTGCCAGCAGTCCTGGGACTACTCAAGAGACCCAGTACTTCG
GGCCAGGCACGCGGCTCCTGGTGCTCGAGGATCTCCGGAatgtgacacccccctaaggtgtctctgttc
gagccaagcaaggccgagatcgccaacaagcagaaggccaccctggtgtgcctggccagaggcttctttcccgatcacgtggag
ctgtcctggtgggtgaatggcaaggaggtgcactctggcgtgtgcaccgaccctcaggcctataaggagtccaactactcttattg
tctgagctcccggctgagagtgtccgccacattctggcacaatcccagaaaccacttcagatgccaggtgcagtttcacggcctgt
ccgaggaggataagtggcctgagggctctccaaagcccgtgacccagaatatcagcgccgaggcatggggaagggcagactgt
ggaatcacctccgcctcttaccagcagggcgtgctgagcgccacaatcctgtatgagatcctgctgggcaaggccaccctgtacg
ccgtgctggtgagcacactggtggtcatggctatggtgaagaggaagaacagccgggccaagcgcagcggctccggcgcaacc
aacttctctctgctgaagcaggcaggcgacgtggaggagaatcctggcccaatgctgactgccagcctgttgagggcagtcatag
cctccatctgtgttgtatccagcatgGCTCAGAAGGTAACTCAAGCGCAGACTGAAATTTCTGTGGT
GGAGAAGGAGGATGTGACCTTGGACTGTGTGTATGAAACCCGTGATACTACTTATTAC
TTATTCTGGTACAAGCAACCACCAAGTGGAGAATTGGTTTTCCTTATTCGTCGGAACTC
TTTTGATGAGCAAAATGAAATAAGTGGTCGGTATTCTTGGAACTTCCAGAAATCCACCA
GTTCCTTCAACTTCACCATCACAGCCTCACAAGTCGTGGACTCAGCAGTATACTTCTGT
GCTCTGAGTCCTGGGGGTAACCAGTTCTATTTTGGGACAGGGACAAGTTTGACGGTCA
TTCCAAATATTCAGAATCCCGAGCCCGCGGtataccagctgaaggacccccggagccaggatagcaccctg
tgcctgttcacagactttgattctcagatcaacgtgcccaagacaatggagagcggcacctttatcacagacaagtgcgtgctgga
catgaaggctatggactctaagagcaatggcgccatcgcctggtccaaccagacctctttcacatgccaggatatctttaaggag
acaaatgccacataccccagctccgacgtgccttgtgatgccaccctgacagagaagagcttcgagacagacatgaatctgaact
ttcagaacctgctggtcatcgtgctgagaatcctgctgctgaaagtggctggcttcaacctgctgatgaccctgcggctgtggagta
gcTAA

Pair 4:

atggactcctggaccttctgctgtgtgtccctttgcatcctggtagcgaagcatacaGATGCTGGAGTTATCCAGTC
ACCCCGCCATGAGGTGACAGAGATGGGACAAGAAGTGACTCTGAGATGTAAACCAATT
TCAGGCCACAACTCCCTTTTCTGGTACAGACAGACCATGATGCGGGGACTGGAGTTGCT
CATTTACTTTAACAACAACGTTCCGATAGATGATTCAGGGATGCCCGAGGATCGATTCT
CAGCTAAGATGCCTAATGCATCATTCTCCACTCTGAAGATCCAGCCCTCAGAACCCAGG
GACTCAGCTGTGTACTTCTGTGCCAGCAGTTTCTCGACCTGTTCGGCTAACTATGGCTA
CACCTTCGGTTCGGGGACCAGGTTAACCGTTGTAGAGGATCTCCGGAatgtgacacccccctaag
gtgtctctgttcgagccaagcaaggccgagatcgccaacaagcagaaggccaccctggtgtgcctggccagaggcttctttcccg
atcacgtggagctgtcctggtgggtgaatggcaaggaggtgcactctggcgtgtgcaccgaccctcaggcctataaggagtccaa
ctactcttattgtctgagctcccggctgagagtgtccgccacattctggcacaatcccagaaaccacttcagatgccaggtgcagtt
tcacggcctgtccgaggaggataagtggcctgagggctctccaaagcccgtgacccagaatatcagcgccgaggcatggggaa
gggcagactgtggaatcacctccgcctcttaccagcagggcgtgctgagcgccacaatcctgtatgagatcctgctgggcaaggc
caccctgtacgccgtgctggtgagcacactggtggtcatggctatggtgaagaggaagaacagccgggccaagcgcagcggct
ccggcgcaaccaacttctctctgctgaagcaggcaggcgacgtggaggagaatcctggcccaatgctgactgccagcctgttga
gggcagtcatagcctccatctgtgttgtatccagcatgGCTCAGAAGGTAACTCAAGCGCAGACTGAAATT
TCTGTGGTGGAGAAGGAGGATGTGACCTTGGACTGTGTGTATGAAACCCGTGATACTA
CTTATTACTTATTCTGGTACAAGCAACCACCAAGTGGAGAATTGGTTTTCCTTATTCGTC
GGAACTCTTTTGATGAGCAAAATGAAATAAGTGGTCGGTATTCTTGGAACTTCCAGAAA
TCCACCAGTTCCTTCAACTTCACCATCACAGCCTCACAAGTCGTGGACTCAGCAGTATA
CTTCTGTGCTCTGAGTCCTGGGGGTAACCAGTTCTATTTTGGGACAGGGACAAGTTTGA
CGGTCATTCCAAATATTCAGAATCCCGAGCCCGCGGtataccagctgaaggacccccggagccaggat
agcaccctgtgcctgttcacagactttgattctcagatcaacgtgcccaagacaatggagagcggcacctttatcacagacaagtg
cgtgctggacatgaaggctatggactctaagagcaatggcgccatcgcctggtccaaccagacctctttcacatgccaggatatct
ttaaggagacaaatgccacataccccagctccgacgtgccttgtgatgccaccctgacagagaagagcttcgagacagacatga
atctgaactttcagaacctgctggtcatcgtgctgagaatcctgctgctgaaagtggctggcttcaacctgctgatgaccctgcggc
tgtggagtagcTAA

Pair 5:

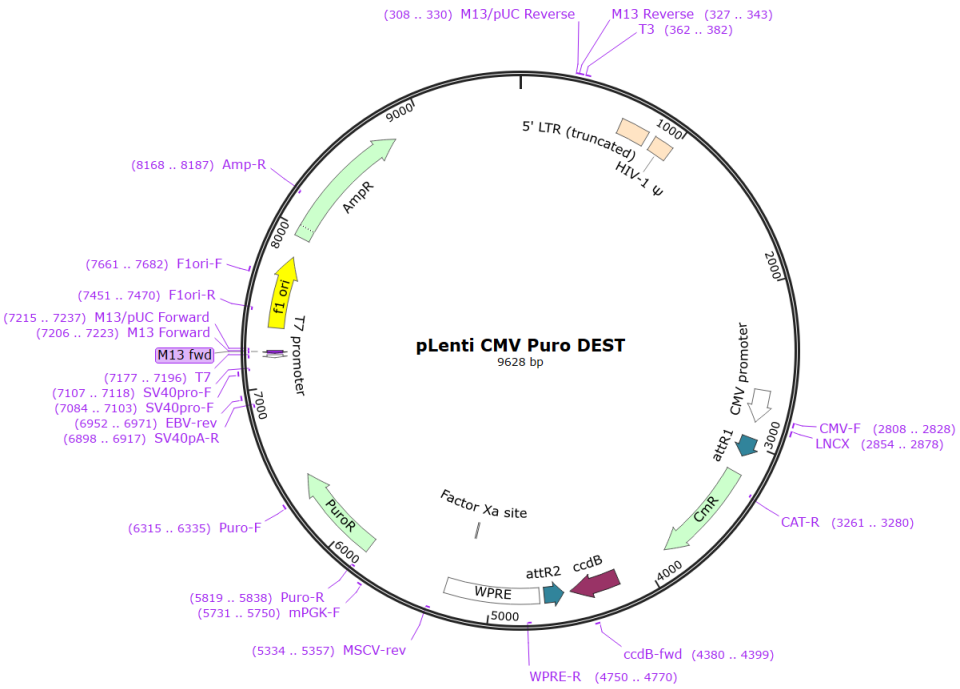
atgctgctgcttctgctgcttctggggccaggtataagcctccttctacctgggagcttggcaggctccgggcttGGTGCTGT
CGTCTCTCAACATCCGAGCAGGGTTATCTGTAAGAGTGGAACCTCTGTGAAGATCGAG
TGCCGTTCCCTGGACTTTCAGGCCACAACTATGTTTTGGTATCGTCAGTTCCCGAAACA
GAGTCTCATGCTGATGGCAACTTCCAATGAGGGCTCCAAGGCCACATATGAGCAAGGC
GTCGAGAAGGACAAGTTTCTCATCAACCATGCAAGCCTGACCTTGTCCACTCTGACAGT
GACCAGTGCCCATCCTGAAGACAGCAGCTTCTACATCTGCAGTGCTAGCGAGGGCTGG
GGCAATCAGCCCCAGCATTTTGGTGATGGGACTCGACTCTCCATCCTAGAGGATCTCCG
GAatgtgacacccccctaaggtgtctctgttcgagccaagcaaggccgagatcgccaacaagcagaaggccaccctggtgtgcc
tggccagaggcttctttcccgatcacgtggagctgtcctggtgggtgaatggcaaggaggtgcactctggcgtgtgcaccgaccct
caggcctataaggagtccaactactcttattgtctgagctcccggctgagagtgtccgccacattctggcacaatcccagaaacca
cttcagatgccaggtgcagtttcacggcctgtccgaggaggataagtggcctgagggctctccaaagcccgtgacccagaatatc
agcgccgaggcatggggaagggcagactgtggaatcacctccgcctcttaccagcagggcgtgctgagcgccacaatcctgtat
gagatcctgctgggcaaggccaccctgtacgccgtgctggtgagcacactggtggtcatggctatggtgaagaggaagaacagc

Pair 5 cntd:

cgggccaagcgcagcggctccggcgcaaccaacttctctctgctgaagcaggcaggcgacgtggaggagaatcctggcccaat
gctgactgccagcctgttgagggcagtcatagcctccatctgtgttgtatccagcatgGCTCAGAAGGTAACTCAAG
CGCAGACTGAAATTTCTGTGGTGGAGAAGGAGGATGTGACCTTGGACTGTGTGTATGA
AACCCGTGATACTACTTATTACTTATTCTGGTACAAGCAACCACCAAGTGGAGAATTGG
TTTTCCTTATTCGTCGGAACTCTTTTGATGAGCAAAATGAAATAAGTGGTCGGTATTCT
TGGAACTTCCAGAAATCCACCAGTTCCTTCAACTTCACCATCACAGCCTCACAAGTCGT
GGACTCAGCAGTATACTTCTGTGCTCTGAGTCCTGGGGGGTAACCAGTTCTATTTTGGGA
CAGGGACAAGTTTGACGGTCATTCCAAATATTCAGAATCCCGAGCCCGCGGtataccagctg
aaggacccccggagccaggatagcaccctgtgcctgttcacagactttgattctcagatcaacgtgcccaagacaatggagagc
ggcacctttatcacagacaagtgcgtgctggacatgaaggctatggactctaagagcaatggcgccatcgcctggtccaaccaga
cctctttcacatgccaggatatctttaaggagacaaatgccacataccccagctccgacgtgccttgtgatgccaccctgacagag
aagagcttcgagacagacatgaatctgaactttcagaacctgctggtcatcgtgctgagaatcctgctgctgaaagtggctggctt
caacctgctgatgaccctgcggctgtggagtagcTAA

APPENDIX J

PLENTI-CMV-PURO VECTOR AND PRIMERS



The pLenti-CMV-Puro vector was used to improve transduction efficiency. TCR αβ pairs were cloned using the attR1/R2 sites. Pairs in the MSGV1 vectors were PCR amplified and flanked with attB1/B2 sites, cloned into the pDONR221 vector, and sequence verified prior to LR reaction into the pLenti-CMV-Puro vector (primer sequences below). Map created with Snapgene.

Forward primers: 20 bp of target gene- ***bold/italic/underlined flanking region***, *italics are the attB1/B2 sites,* the **bold letter is Shine-Delgarno**, the grey highlight is the Kozak region, underlined is the gene with the start codon

Pair 1
5'- ***G GGG*** *ACA AGT TTG TAC AAA AAA GCA GGC TTC* **GAA GGA GAT AGA** ACC ATG GGC ACC AGG CTC CTC TTC -3'
Pair 2
5'-***G GGG*** *ACA AGT TTG TAC AAA AAA GCA GGC TTC* **GAA GGA GAT AGA** ACC ATG GGC ACC AGG CTC CTC -3'
Pair 3
5'-***G GGG*** *ACA AGT TTG TAC AAA AAA GCA GGC TTC* **GAA GGA GAT AGA** ACC ATG GGC CCC CAG CTC CTT -3'
Pair 4
5'-***G GGG*** *ACA AGT TTG TAC AAA AAA GCA GGC TTC* **GAA GGA GAT AGA** ACC ATG GAC TCC TGG ACC TTC -3'
Pair 5
5'-***G GGG*** *ACA AGT TTG TAC AAA AAA GCA GGC TTC* **GAA GGA GAT AGA** ACC ATG GGG ATG CTG CTG C -3'

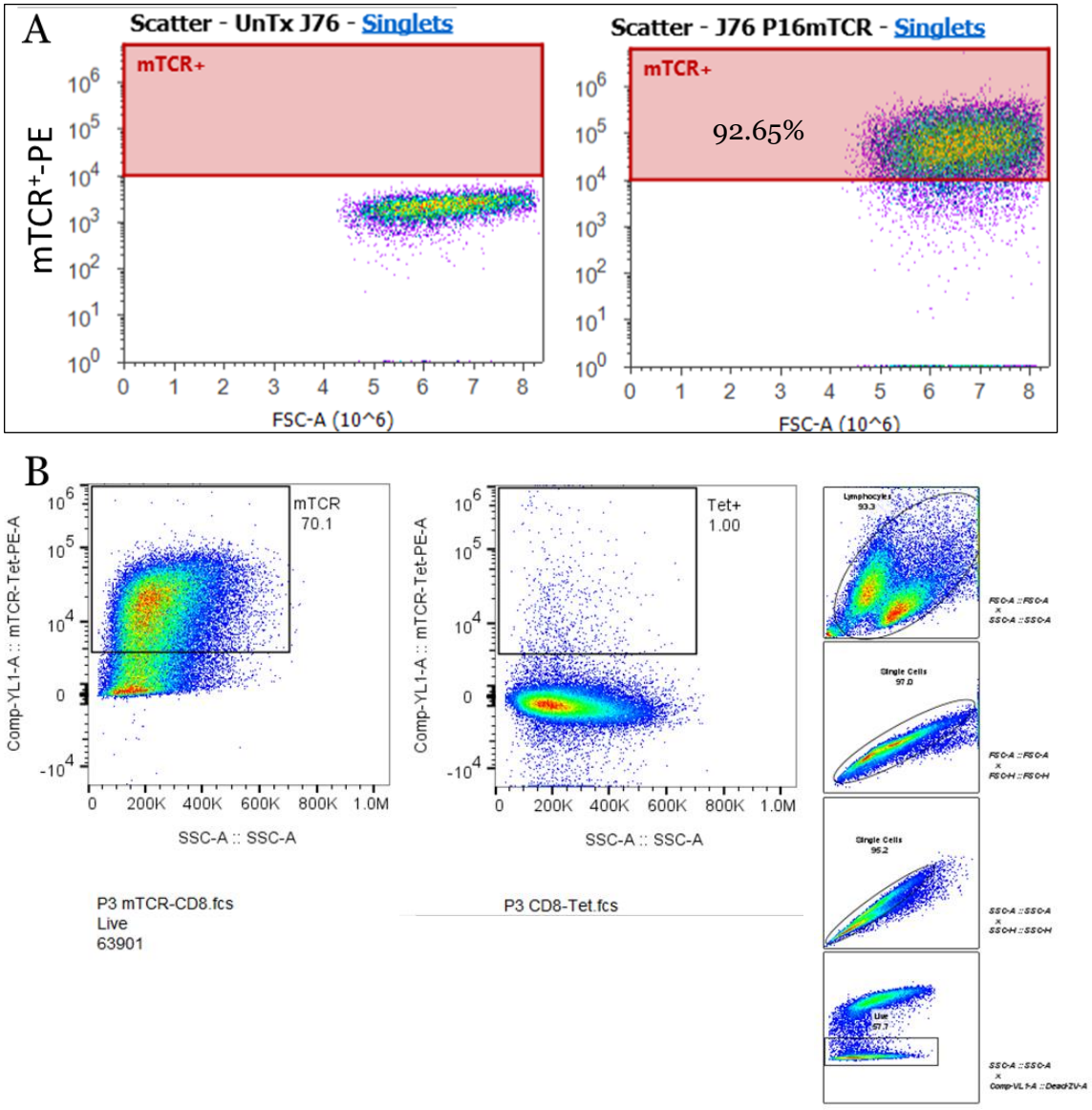
Reverse primer is the same for all constructs since it complements a constant region that is the same for all TCR pairs.

Reverse primer with 20 bp of target gene- ***bold/italic/underlined is suggested flanking region of 5'***, *italics are the attB1/B2sites,* underlined is the gene with two stop codons:

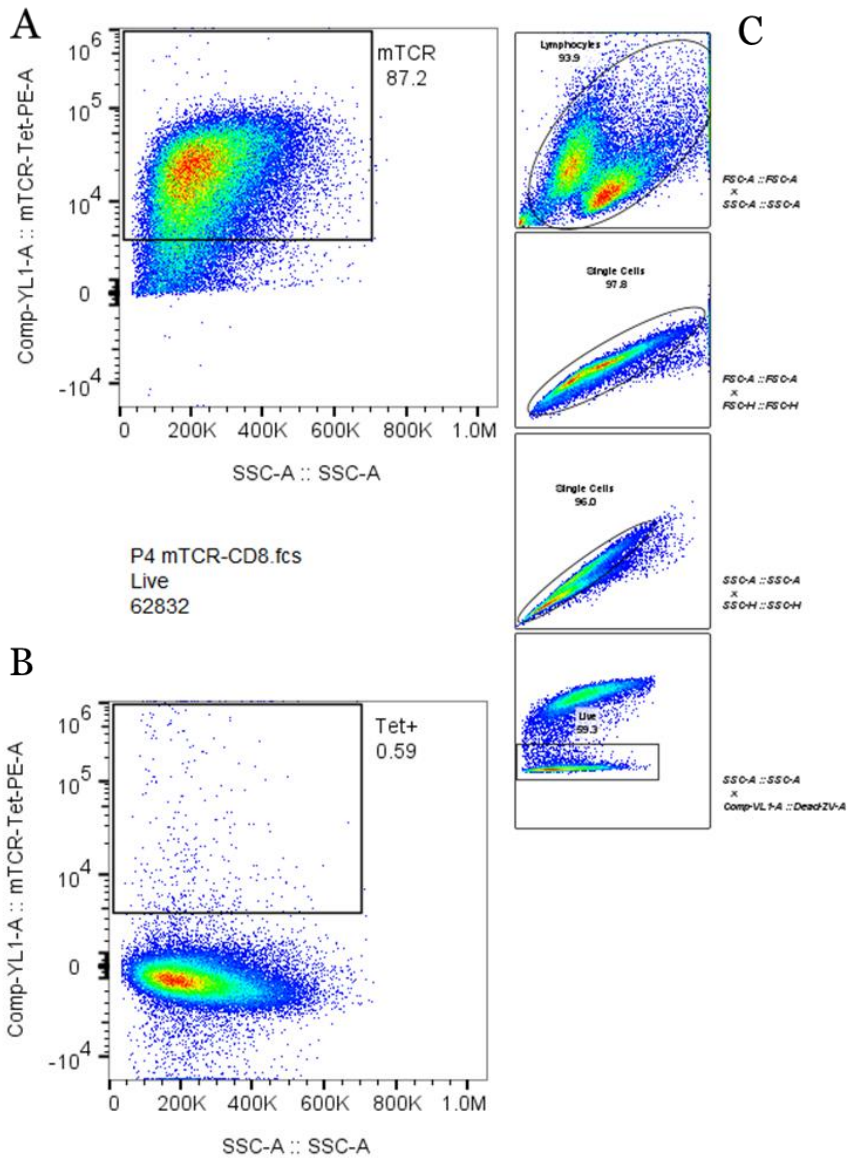
5'-***GGG G****AC CAC TTT GTA CAA GAA AGC TGG GTC CTA* TCA TTA GCT ACT CCA CAG CCG CAG-3'

APPENDIX K

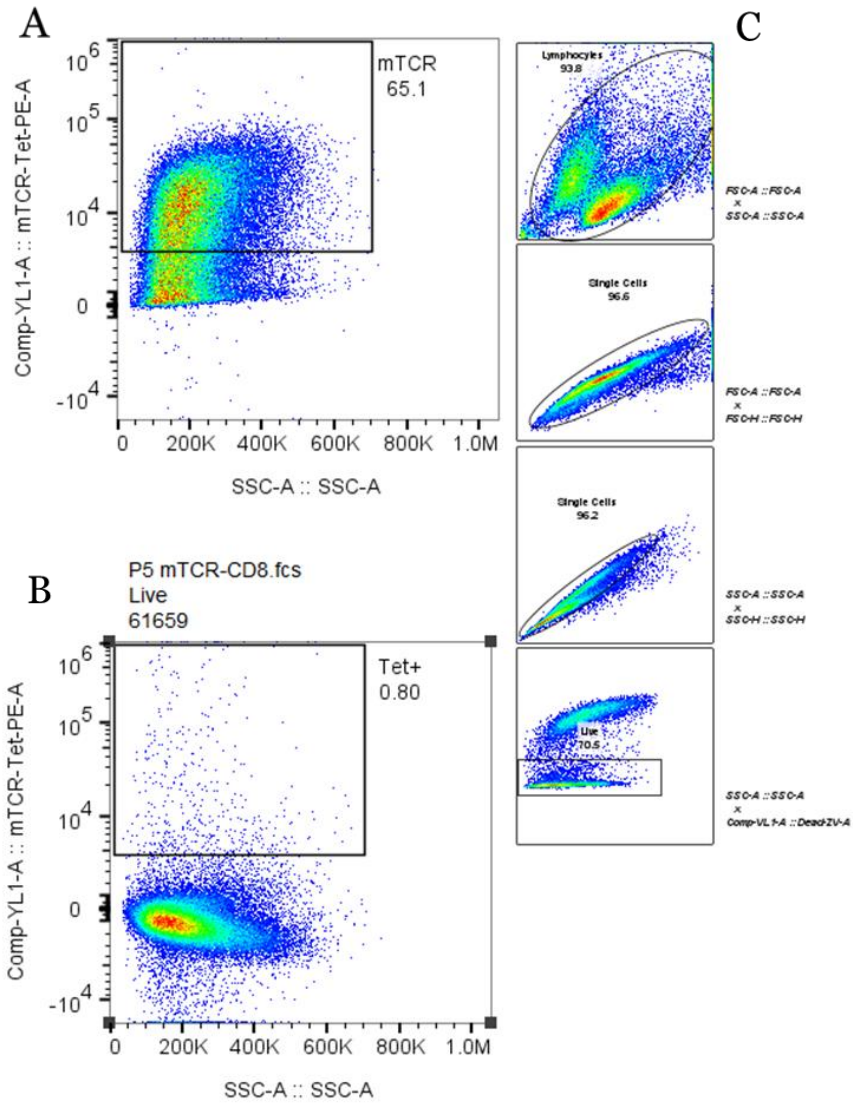
FLOW CYTOMETRY PANELS OF TCR EXPRESSION IN MSGV1 AND PLENTI T CELL LINES



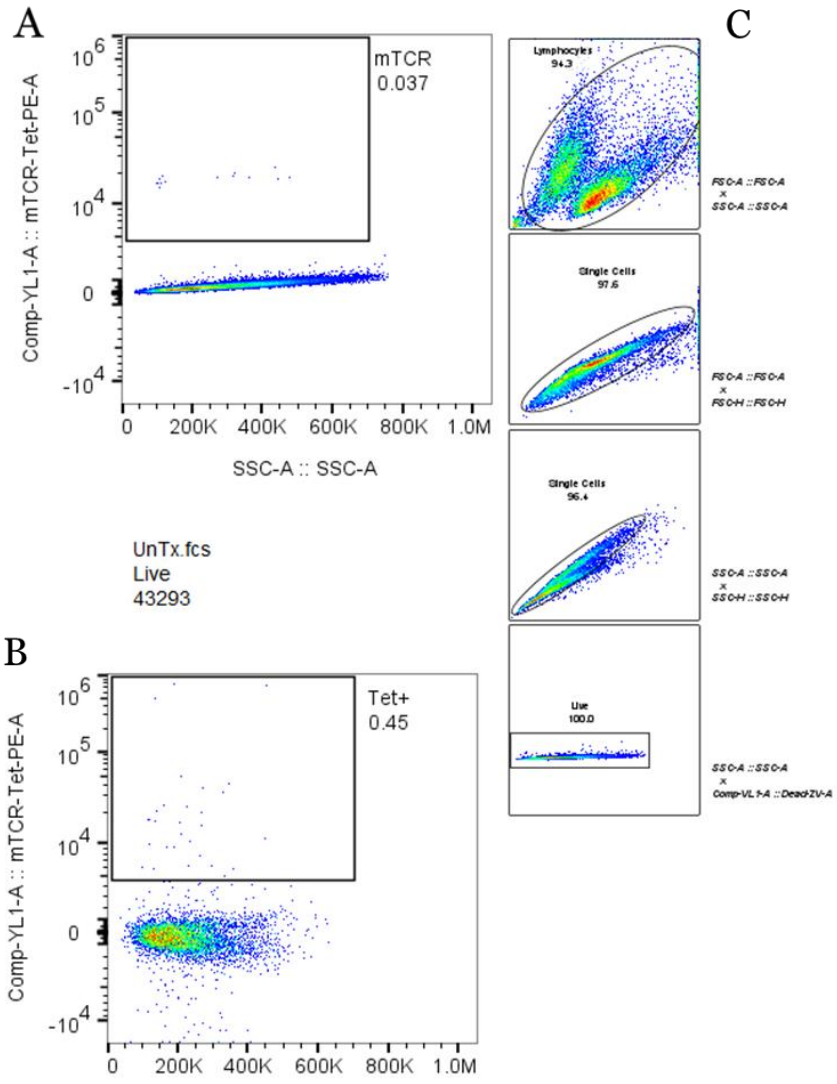
(A). mTCR^{+} expression of J76-CD8^{+} T cells transduced with Pair #2 of the KAYN-reactive TCR. The left panel are untransduced J76-CD8^{+} T cells. The right panel are J76-CD8^{+} T cells 5 days post-transduction. (B). The Left Panel are J76-CD8^{+} T cells transduced with the KAY^{N}-TCR Pair #3 pLenti construct. The center panel are the Tet^{+} CD8^{+} J76 T cells. Panels on the right correspond to the gated ancestry (All cells, single cells FSC-forward scatter and SSC-Side scatter, and live cells).



Top left panel (A)J76-CD8⁺ T cells transduced with the KAY^{N}-TCR Pair #4 pLenti construct. Bottom panel (B) Tet+ CD8+ J76 T cells. Panels on the right (C) gated ancestry (All cells, single cells FSC-forward scatter and SSC-Side scatter and live cells).



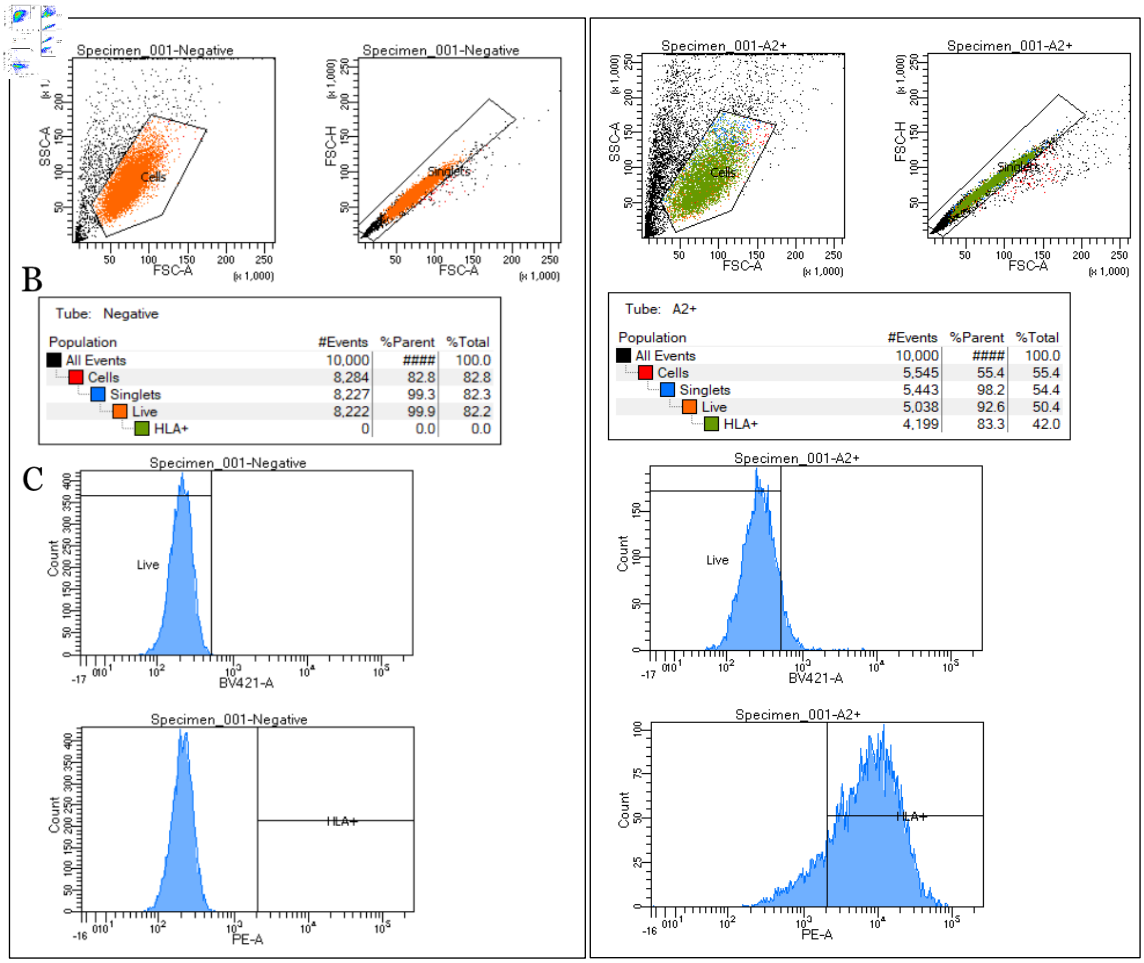
Top left panel (A)J76-CD8^{+} T cells transduced with the KAY^{N}-TCR Pair #5 pLenti construct. Bottom panel (B) Tet+ CD8+ J76 T cells. Panels on the right (C) gated ancestry (All cells, single cells FSC-forward scatter and SSC-Side scatter and live cells).



Untransduced cells. Top left panel (A) J76-CD8⁺ T cells stained with anti-mTCR-PE. Bottom panel (B) Untransduced cells stained with KAY^{N}-Tetramer-PE. Panels on the right (C) gated ancestry (All cells, single cells FSC-forward scatter and SSC-Side scatter and live cells).

APPENDIX L

FLOW CYTOMETRY PANELS OF HLA EXPRESSION OF K562 ARTIFICIAL ANTIGEN-PRESENTING STABLE CELL LINES



Left panel: (A): Flow cytometry panel of untransduced K562 cells gated forward (FSC) and side scatter (SSC). (B): Cell counts by gate. (C): Top and bottom panels are histograms of live and HLA^{+} cells. Right panel: (A): Flow cytometry panel of HLA-A*02:01 pLenti-CMV-Puro transduced K562 cells gated forward (FSC) and side scatter (SSC). (B): Cell counts by gate. (C): Top and bottom panels are histograms of live and HLA+ cells. Cells were bulk sorted, selected with puromycin for 5 days, and aliquoted into 1x10^{6} per mL stocks for further use.

APPENDIX M

RIGHTS AND PERMISSIONS

Chapter 1

Figure 1. Adapted from an open-source template originally published by: Nguyen, A. T; Szeto, C. and Gras, S. The pockets guide to HLA class I molecules. Biochem. Soc. Trans. 2021. 49(5): 2319-2331.

Table 1. Adapted from an open-source template originally published by: Sidney, J., et al., *HLA class I supertypes: a revised and updated classification*. BMC Immunol, 2008. **9**: p. 1

Chapter 4

Figure 23. Adapted from an open-source template originally published by: Laydon, D.J., C.R. Bangham, and B. Asquith, *Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach*. Philos Trans R Soc Lond B Biol Sci, 2015. **370**(1675).