



SARS-CoV-2 mutations affect antigen processing by the proteasome to alter CD8⁺ T cell responses

Dannielle Wellington^{a,b,1,*}, Zixi Yin^{a,b,1}, Zhanru Yu^{a,c,1}, Raphael Heilig^c, Simon Davis^{a,c}, Roman Fischer^{a,c}, Suet Ling Felce^{a,d}, Elie Antoun^{a,d}, Philip Hublitz^e, Ryan Beveridge^f, Danning Dong^{a,b}, Guihai Liu^{a,b}, Xuan Yao^{a,b}, Yanchun Peng^{a,b}, Benedikt M. Kessler^{a,c,1}, Tao Dong^{a,b,1,**}

^a Chinese Academy of Medical Sciences (CAMS) Oxford Institute, Nuffield Department of Medicine, Oxford University, Oxford, OX3 7FZ, UK

^b MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, Oxford University, Oxford, OX3 9DS, UK

^c Target Discovery Institute, Centre for Medicines Discovery, Nuffield Department of Medicine, Oxford University, Oxford, OX3 7FZ, UK

^d Wellcome Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK

^e Genome Engineering Facility, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, Oxford University, Oxford, OX3 9DS, UK

^f Virus Screening Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, UK

ABSTRACT

Mutations within viral epitopes can result in escape from T cells, but the contribution of mutations in flanking regions of epitopes in SARS-CoV-2 has not been investigated. Focusing on two SARS-CoV-2 nucleoprotein CD8⁺ epitopes, we investigated the contribution of these flanking mutations to proteasomal processing and T cell activation. We found decreased NP₉₋₁₇-B*27:05 CD8⁺ T cell responses to the NP-Q7K mutation, likely due to a lack of efficient epitope production by the proteasome, suggesting immune escape caused by this mutation. In contrast, NP-P6L and NP-D103 N/Y mutations flanking the NP₉₋₁₇-B*27:05 and NP₁₀₅₋₁₁₃-B*07:02 epitopes, respectively, increased CD8⁺ T cell responses associated with enhanced epitope production by the proteasome. Our results provide evidence that SARS-CoV-2 mutations outside the epitope could have a significant impact on proteasomal processing, either contributing to T cell escape or enhancement that may be exploited for future vaccine design.

1. Introduction

Defining CD8⁺ T cell responses to SARS-CoV-2 has been a priority since the virus first emerged in 2019 as these responses are instrumental in clearing viral infections [1]. Multiple immunodominant CD8⁺ T cell epitopes have been identified [2–7]. Generation of CD8 epitopes initially requires antigen processing through the proteasome, a multi-subunit cytosolic machinery that can degrade intracellular proteins following ubiquitination. The proteasome is composed of a catalytic core particle (20s) and one or two terminal 19s regulatory particles. The 20s catalytic core is capable of cleaving at the C-terminal side of acidic, basic and hydrophobic amino acid residues [8–10]. As part of this proteolytic process, a small fraction of peptide fragments will contain correctly cleaved C-termini of

* Corresponding author. Chinese Academy of Medical Sciences (CAMS) Oxford Institute, Nuffield Department of Medicine, Oxford University, Oxford OX3 7FZ, UK.

** Corresponding author. Chinese Academy of Medical Sciences (CAMS) Oxford Institute, Nuffield Department of Medicine, Oxford University, Oxford, OX3 7FZ, UK.

E-mail addresses: dannielle.wellington@rdm.ox.ac.uk (D. Wellington), tao.dong@ndm.ox.ac.uk (T. Dong).

¹ These authors contributed equally to this work.

<https://doi.org/10.1016/j.heliyon.2023.e20076>

Received 6 September 2022; Received in revised form 22 August 2023; Accepted 11 September 2023

Available online 14 September 2023

2405-8440/© 2023 Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

CD8 epitopes, whereas the majority contain extended N-termini that require further trimming. This can occur in the cytoplasm by various peptidases or, more commonly, in the endoplasmic reticulum (ER) following translocation of peptides from the cytosol via the transporter associated with antigen processing complex TAP1/TAP2 [11,12].

An essential peptidase responsible for trimming peptides in the ER is ERAP1; an aminopeptidase optimised to generate peptides of 8–10 amino acids for binding to MHC Class I molecules [13]. ERAP1 activity varies depending on the N-terminal amino acid with little ability to cleave after proline residues and high ability to cleave after hydrophobic amino acids [14]. Another related aminopeptidase, ERAP2, can also help to shape the MHC Class I peptide repertoire, but has a slightly different trimming capacity preferring positively charged amino acids [15].

Once cleaved to optimal length, peptides will bind to MHC Class I molecules and those with sequences that bind strongly into the

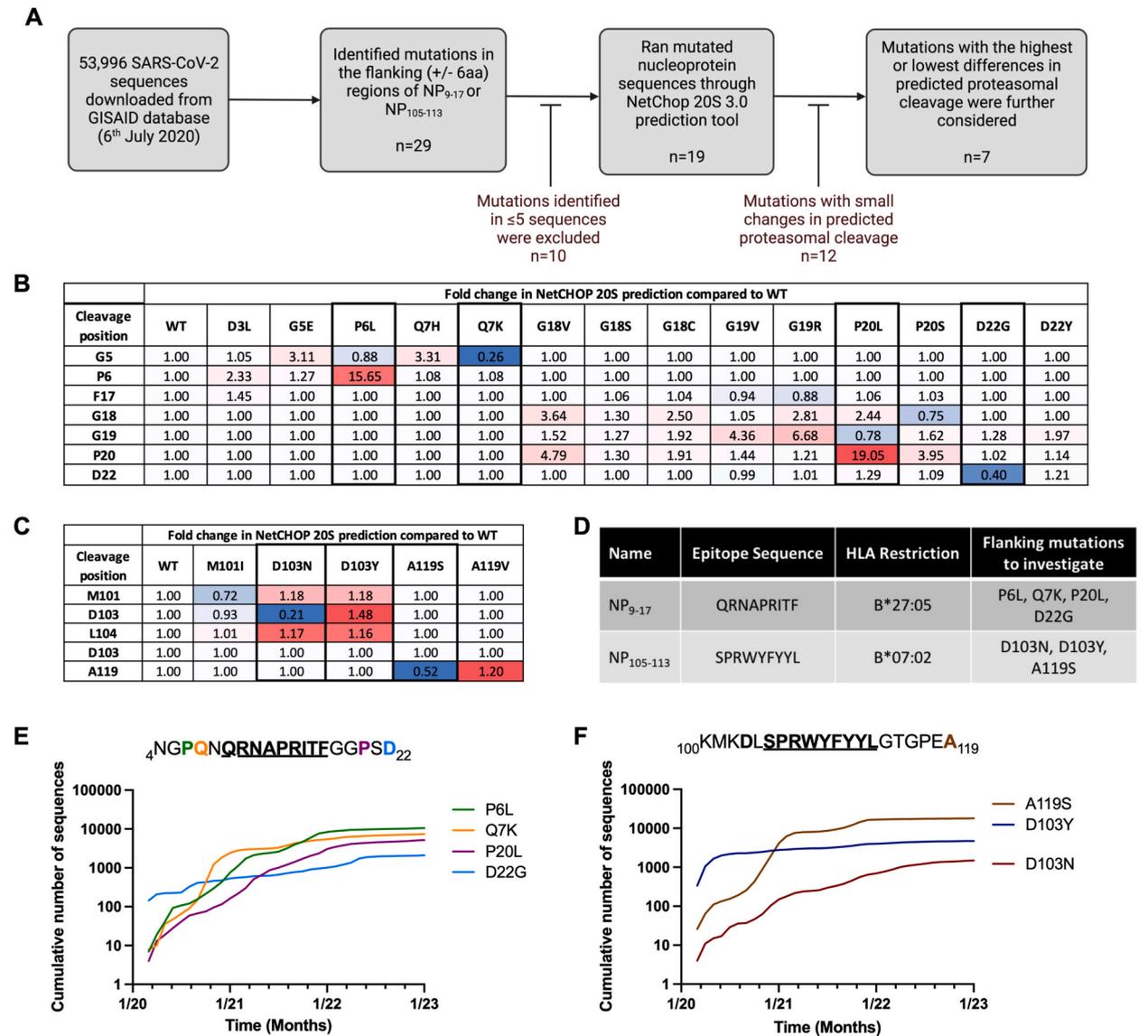


Fig. 1. Identification of mutations in SARS-CoV-2 CD8⁺ T cell nucleoprotein epitope flanking regions. (A) 53,996 SARS-CoV-2 sequences were downloaded from the GISAID database on July 6, 2020. Mutations in the flanking regions (± 6 amino acids) were assessed either side of identified CD8 nucleoprotein epitopes (NP₉₋₁₇/NP₁₀₅₋₁₁₃) and analysed the potential effect these mutations may have on antigen processing by the proteasome. (B) Heatmap results from NetChop 20s prediction tool for the effect of mutations on proteasomal cleavage of NP₉₋₁₇ region. Red = increase in cleavage, blue = decrease. (C) Heatmap results from NetChop 20s prediction tool for the effect of mutations on proteasomal cleavage of NP₁₀₅₋₁₁₃ region. Red = increase in cleavage, blue = decrease. (D) Mutations taken forward for further investigation. (E) Frequency of mutations for NP₄₋₂₂ region expressed as cumulative number of sequences with each mutation since January 2020. (F) Frequency of mutations for NP₁₀₀₋₁₁₉ region expressed as cumulative number of sequences with each mutation since January 2020. Data downloaded from COVID-19 Viral Genome Analysis Pipeline.

peptide-binding groove, so-called high affinity peptides, will induce transport of peptide-MHC complexes to the cell surface for activation of CD8⁺ T cells [16]. Thus, the sequence of the peptides is important for ensuring strong binding to MHC, and mutations within these sequences can alter the affinity of this interaction. Using exogenous peptide assays, the effect of mutations within epitope sequences have been previously shown to lead to escape from viral T cell responses in various settings [17–19] including SARS-CoV-2 [20].

In addition, components of the antigen processing pathway have preferences for amino acid sequences for their optimal processing as discussed above. Indeed, previous studies [21] looking into the effect of viral mutations in the flanking regions of epitopes on antigen processing of HIV-1 found evidence of disruption of proteasomal cleavage [22] and endoplasmic reticulum aminopeptidase 1 (ERAP1) trimming within the ER [23].

Traditionally, single-stranded RNA viruses have a high propensity for mutation due to a lack of proof-reading ability by the RNA-dependent RNA polymerases that replicate their genomes [24]. However, coronaviruses have adapted proofreading capabilities that reduce their mutation rate [25]. Although several variants of concern have still emerged since the start of the pandemic. Currently, the effects of SARS-CoV-2 epitope-flanking mutations on viral epitope antigen processing are unknown. Our aim of this proof-of-concept study was to identify whether mutations are occurring in these regions with possible effects on CD8⁺ T cell activation.

2. Results

2.1. Identification of mutations in flanking regions of SARS-CoV-2 nucleoprotein CD8 epitopes

Focusing on two immunodominant SARS-CoV-2 nucleoprotein (NP) epitopes previously identified, NP₉₋₁₇ and NP₁₀₅₋₁₁₃ [2,26], we searched sequences derived from the GISAID database (www.gisaid.org; July 6, 2020 (EPI_SET_230323 zm) for mutations within six amino acids either side of these epitopes through sequence alignment with the originally identified Wuhan sequence (hCoV-19/Wuhan/WIV04/2019 (WIV04) (EPI_ISL_402,124)), which yielded 29 mutations (Fig. 1A). Mutations found in five or fewer sequences (<10%) were excluded from further investigation to eliminate mutations that were due to sequencing artefacts, resulting in 19 mutations. We subjected these sequences through the IEDB NetChop 20S 3.0 prediction tool for proteasomal cleavage [27] and compared cleavage against wild-type (WT) NP sequence derived from the viral strain originally isolated in Wuhan, China (Fig. 1B–C). The mutations that showed the greatest predicted differences in proteasomal cleavage as compared to WT NP sequence for each epitope were selected for further investigation (n = 7) (Fig. 1D). While the frequency of these mutations was extremely low at the time of analysis, they have been steadily rising since July 2020 (Fig. 1E–F). The frequency of each mutation compared to the vast number of

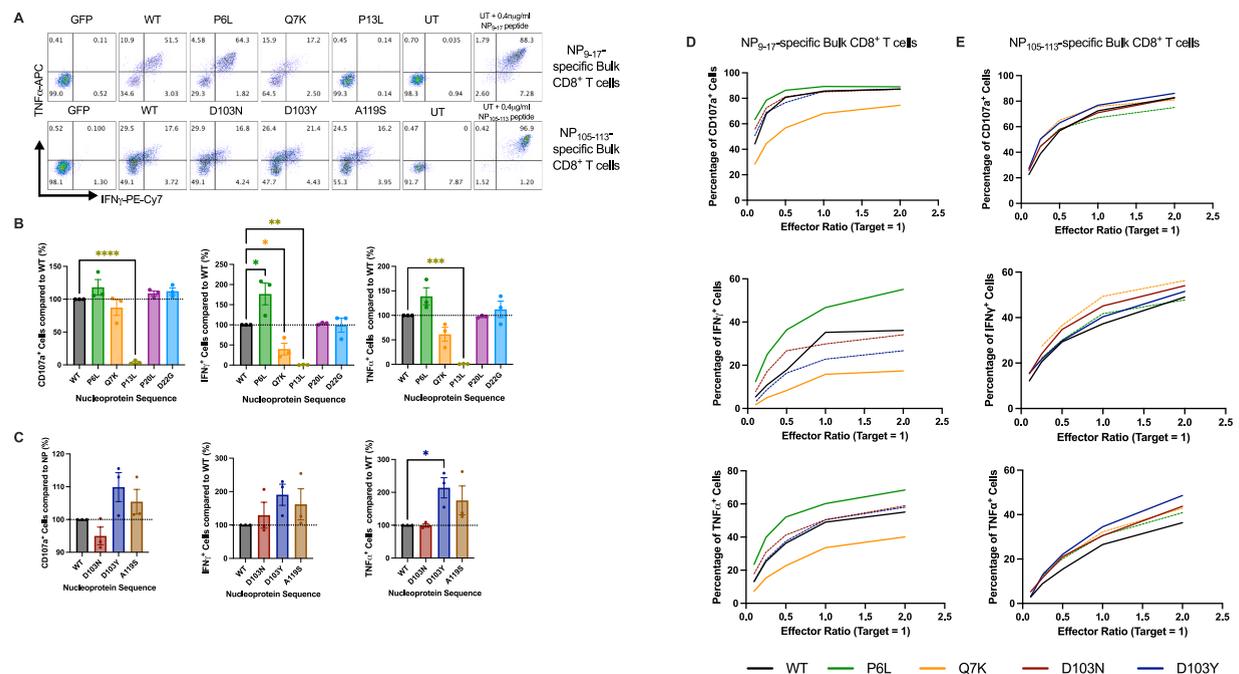


Fig. 2. Mutations in flanking regions of epitopes alter CD8⁺ T Cell responses. (A) Intracellular cytokine staining of NP₉₋₁₇ or NP₁₀₅₋₁₁₃ epitope-specific bulk CD8⁺ T cells following incubation with GFP- or NP-transduced B cells. NP₉₋₁₇ or NP₁₀₅₋₁₁₃ peptide-pulsed B cells are also shown. (B) CD8⁺ T Cell response to mutations flanking NP₉₋₁₇ were analysed by flow cytometry at an E:T ratio of 2:1. (C) CD8⁺ T Cell responses to mutations flanking NP₁₀₅₋₁₁₃ were analysed by flow cytometry at an E:T ratio of 2:1. (D–E) Selected mutations were investigated in more detail by altering the E:T ratio from 2:1 to 0.1:1. Data is presented as mean ± SEM and was analysed by one-way ANOVA and Dunnett’s multiple comparisons test. **p* < 0.05, ***p* < 0.01, ****p* < 0.001, *****p* < 0.0001.

sequences now available shows that these mutations are still relatively rare but 4 out of 7 mutations have increased in frequency across the 2 years studied (Supplementary Table 1). Up until March 2022, they have not appeared in any variants of interest or concern identified by the WHO.

2.2. Unaltered HLA expression in SARS-CoV-2 NP-transduced B cells

Having identified mutations of interest, we next created stable B cell lines expressing SARS-CoV-2 WT or mutated NP through lentiviral transduction. B cells from patient C-COV19-005 isolated and immortalised as previously described in Ref. [2] were chosen for initial experiments as they expressed both HLA-B*27:05 and HLA-B*07:02, the correct MHC class I restriction for NP₉₋₁₇ and NP₁₀₅₋₁₁₃, respectively (full HLA typing for donor C-COV19-005 is available in the methods section). Each NP construct also contained a GFP gene separated by a viral self-cleavage T2A site, resulting in one long transcript but two translated proteins, GFP and NP. NP-transduced cells showed no significant differences in GFP, SARS-CoV-2 NP or HLA expression (Supplementary Fig. 1).

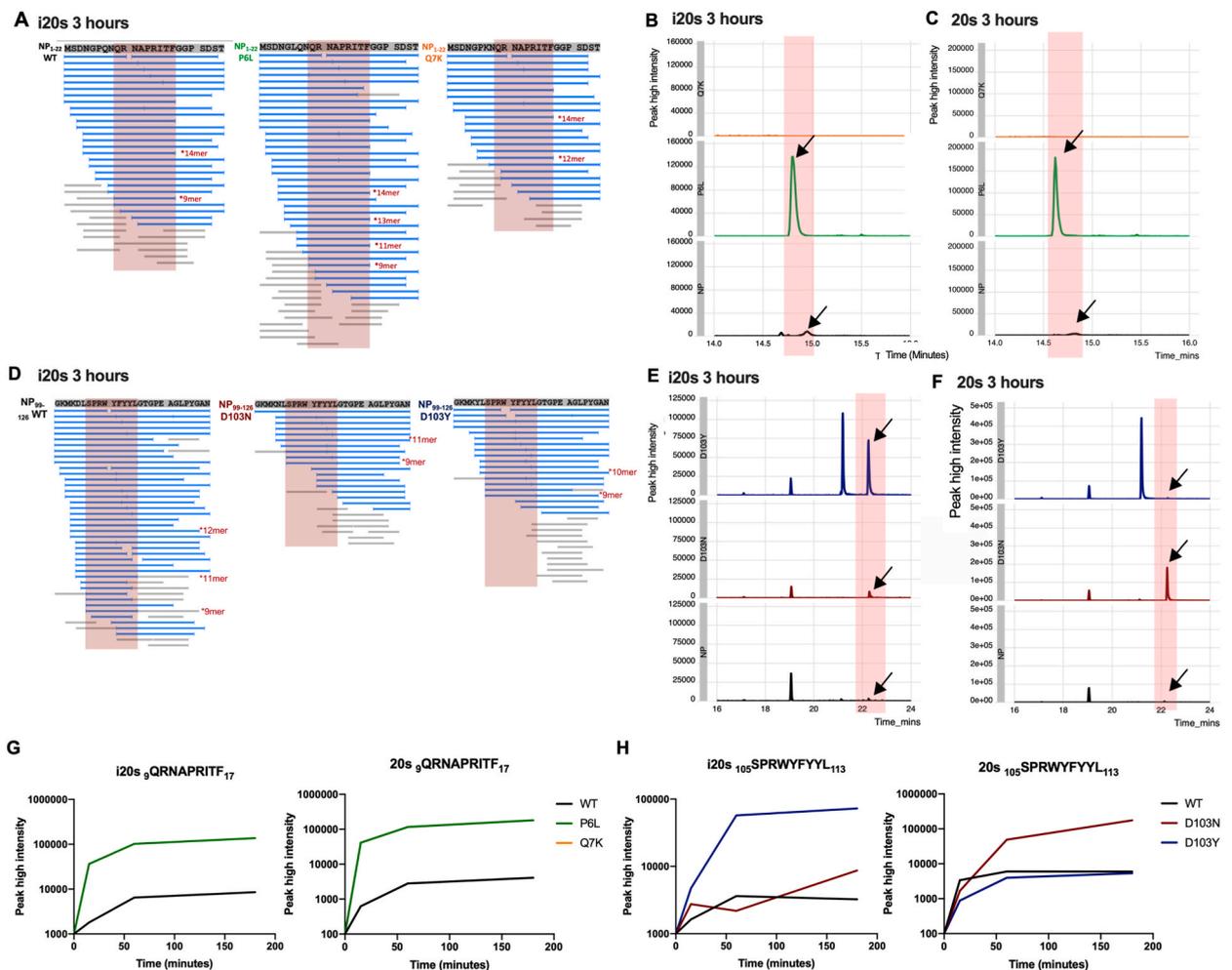


Fig. 3. Mutations in flanking regions affect proteasomal digestion and epitope generation of SARS-CoV-2 CD8⁺ T cell epitopes. (A) PEAKS data showing the products of NP₁₋₂₂ peptide digested with immunoproteasome for 3 h. Correctly C-terminally cleaved peptides are labelled for 9-14mer peptides. (B) Chromatograms showing generation of NP₉₋₁₇ epitope after 3-h incubation of NP₁₋₂₂ peptide with immunoproteasome provide peak high intensity values. (C) Chromatograms showing generation of NP₉₋₁₇ epitope after 3-h incubation of NP₁₋₂₂ peptide with constitutive proteasome provide peak high intensity values. (D) PEAKS data showing the products of NP₉₉₋₁₂₆ peptide digested with immunoproteasome for 3 h. Correctly C-terminally cleaved peptides are labelled for 9-14mer peptides. (E) Chromatograms showing generation of NP₁₀₅₋₁₁₃ epitope after 3-h incubation of NP₉₉₋₁₂₆ peptide with immunoproteasome provide peak high intensity values. The additional peaks correspond to NP₁₀₃₋₁₁₀ and NP₁₁₂₋₁₂₄. (F) Chromatograms showing generation of NP₁₀₅₋₁₁₃ epitope after 3-h incubation of NP₉₉₋₁₂₆ peptide with constitutive proteasome provide peak high intensity values. The additional peaks correspond to NP₁₀₃₋₁₁₀ and NP₁₁₂₋₁₂₄. (G) Time-course of NP₉₋₁₇ epitope generation from NP₁₋₂₂ long peptide with variants digested with immuno- or constitutive proteasome. (H) Time-course of NP₁₀₅₋₁₁₃ epitope generation from NP₉₉₋₁₂₆ long peptide with variants digested with immuno- or constitutive proteasome.

2.3. Flanking mutations alter CD8⁺ T cell responses

Our NP-transduced B cell system allowed us to evaluate the effect of one amino acid mutation on T cell responses to NP. We recognise that this is not a complete surrogate for viral infection as NP protein expression in transduced cells may be greater than that of a SARS-CoV-2 infection. However, NP is the most abundant viral protein at all stages of the infection cycle, especially in the initial stages [28,29], and a recent study showed that comparable levels of NP could be detected following virus infection as compared to transfection of HEK293T cells [30].

Using NP-transduced cell lines as target cells, we co-cultured with bulk CD8⁺ T cells-specific for NP₉₋₁₇ or NP₁₀₅₋₁₁₃ (Supplementary Fig. 2) for 4 h prior to analysis of CD107a, IFN γ and TNF α expression (Fig. 2A). To show the specificity of our bulk T cell lines, we show a peptide-pulsed control for each T cell line. We included the mutation NP-P13L as a control in our experiments as previous work had shown this to be a CD8⁺ T cell escape mutant within the epitope NP₉₋₁₇ [20]. This mutation has subsequently been found within the Omicron variant of SARS-CoV-2. As expected, the NP-P13L mutation was shown to lead to complete escape from all NP₉₋₁₇ CD8⁺ T cell responses measured when compared to WT NP (Fig. 2B).

Interestingly, we also saw significant differences in T cell responses to several flanking mutations (Fig. 2B–C). NP-Q7K significantly decreased IFN γ ⁺ NP₉₋₁₇-specific CD8⁺ T cells ($p = 0.0448$). In contrast, NP-P6L was shown to significantly increase the proportion of IFN γ ⁺ NP₉₋₁₇-specific T cells ($p = 0.0107$), while NP-D103Y significantly increased TNF α ⁺ NP₁₀₅₋₁₁₃-specific T cells ($p = 0.0424$).

Addressing the mutations that showed significant differences, in addition to NP-D103 N as it affects the same amino acid position as NP-D103Y, we investigated how these mutations modulate CD8⁺ T cell responses in more detail. The previous experiment was performed at an effector (CD8⁺ T Cell) to target (NP-transduced B cell) ratio (E:T) of 2:1. We next used a range of E:T from 2:1 to 0.1:1 to expose differences in CD8⁺ T cell responses (Fig. 2D–E). For the mutations flanking NP₉₋₁₇, there was clear separation between the results for wild type (WT) vs mutant, with NP-P6L always showing higher and NP-Q7K showing lower T cell responses. For NP₁₀₅₋₁₁₃, the differences were more subtle, but there is separation between WT and mutant proteins. Taken together, these results clearly demonstrate that flanking mutations, particularly N-terminal mutants, can significantly alter CD8⁺ T cell responses, presumably by altering the antigen load of the epitopes on the cell surface.

2.4. Flanking mutations alter epitope generation by the proteasome

CD8⁺ T cell epitopes are generated through degradation of proteins by the proteasome, machinery that selectively cleaves after certain amino acid residues. Therefore, we investigated whether our flanking mutations alter proteasomal digestion of NP epitopes. To do this, we digested long peptide precursors (NP₁₋₂₂ and NP₉₉₋₁₂₆), containing the epitopes NP₉₋₁₇ and NP₁₀₅₋₁₁₃ and either WT or mutated flanking regions with constitutive (20S) or immunoproteasome (i20S) for up to 3 h *in vitro*, followed by analysis of the digestion products via quantitative mass spectrometry (MS).

The immunoproteasome is constitutively expressed in haematopoietic cells or induced in response to IFN γ or TNF α and can produce

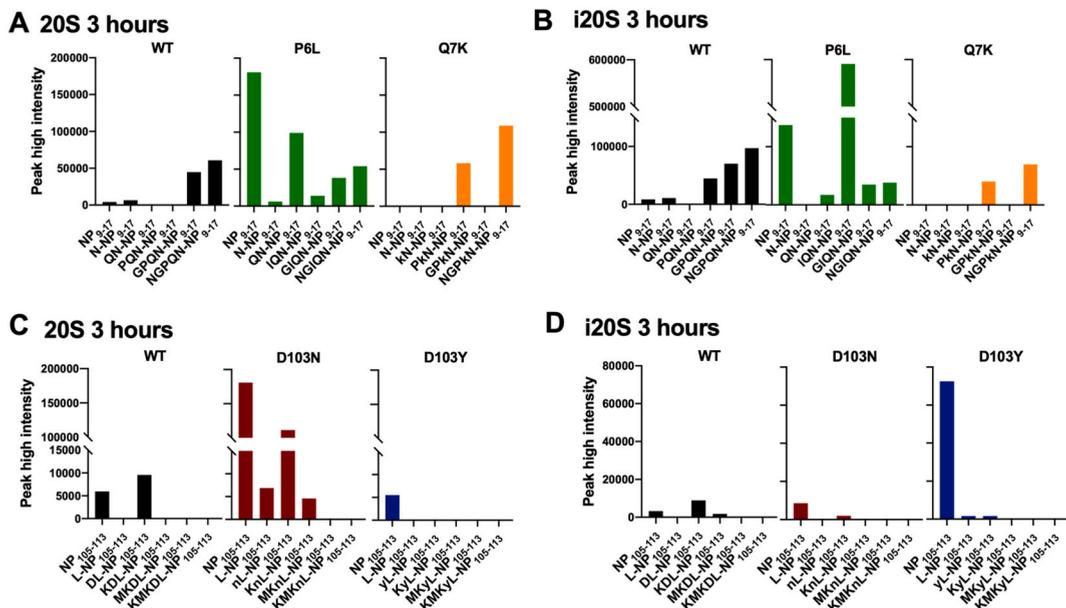


Fig. 4. Flanking mutations alter N-terminally extended proteasomal products. (A) The generation of 9–14mers (9mer = NP₉₋₁₇) from NP₁₋₂₂ peptide following digestion with constitutive proteasome for 3 h. (B) The generation of 9–14mers (9mer = NP₉₋₁₇) from NP₁₋₂₂ peptide following digestion with immunoproteasome for 3 h. (C) The generation of 9–14mers (9mer = NP₁₀₅₋₁₁₃) from NP₉₉₋₁₂₆ peptide following digestion with constitutive proteasome for 3 h. (D) The generation of 9–14mers (9mer = NP₁₀₅₋₁₁₃) from NP₉₉₋₁₂₆ peptide following digestion with immunoproteasome for 3 h.

antigenic peptides at a faster rate than the constitutive proteasome [31,32]. As a haematopoietic cell line, we expected that our B cell lines would express a combination of 20s and i20s proteasomes, which was confirmed through detection of constitutive proteasome (PSMB6) and immunoproteasome (LMP2, LMP7) subunits (Supplementary Fig. 3).

The products of NP₁₋₂₂ with a 3-h immunoproteasome digestion showed a distinct pattern of products for NP-WT, NP-P6L and NP-Q7K (Fig. 3A). Strikingly, for both NP₉₋₁₇ and NP₁₀₅₋₁₁₃ epitopes, proteasomal cleavage alone was sufficient to produce an abundance of correctly C-terminally cleaved peptides (Fig. 3). Additionally, while N-terminally extended versions were also present, the complete nine amino acid (9mer) epitopes were identified from proteasomal cleavage alone. This is unusual [33,34] and warrants further investigation to confirm whether these 9mer peptides can be transported into the ER through TAP for direct loading on HLA Class I molecules, or whether these peptides are not immunogenic at all. In either case, the effect of mutations on proteasomal cleavage is clearly demonstrated.

Using the peak-high intensity values from extracted ion chromatograms as shown in Fig. 3B–C, we were able to get a relative quantitative value for each peptide over a time-course of 3 h. Across this 3-h time-course, we observed higher NP₉₋₁₇ epitope generation with NP-P6L as compared to WT sequences, which was consistent with both immunoproteasome and constitutive proteasome digestions (Fig. 3G). We were unable to detect NP₉₋₁₇ epitope from NP-Q7K mutated protein sequences with either proteasome (Fig. 3A–C, 3G). We identified proteasomal differences in the response to mutated proteins in the NP₉₉₋₁₂₆ peptide, with NP-D103 N increasing NP₁₀₅₋₁₁₃ epitope generation by the constitutive proteasome but having an insignificant effect with the immunoproteasome. In contrast, NP-D103Y enhanced NP₁₀₅₋₁₁₃ epitope generation by immunoproteasome digestion (Fig. 3D–F, 3H).

2.5. Flanking mutations alter N-terminally extended proteasomal products

In addition to differences in 9mer epitope generation by the proteasome, we found that mutations also affected generation of N-terminally extended epitope precursors (Fig. 4). This appears to be particularly prevalent for the mutations NP-P6L (Fig. 4A–B) and NP-D103 N (Fig. 4C). We also observed differences dependent on proteasome species (Fig. 4C–D). N-terminally extended precursors are most commonly trimmed by endoplasmic reticulum aminopeptidase 1 (ERAP1) in the ER to the optimal length for HLA Class I binding [35], confirmed to be nine amino acids for NP₉₋₁₇ and NP₁₀₅₋₁₁₃ [2]. The requirement for further processing of epitopes can alter the efficiency of antigen processing and presentation potentially affecting antigen load at the cell surface.

3. Discussion

For SARS-CoV-2 infection, it has been demonstrated that mutations within epitope sequences can lead to escape from T cell recognition [20], presumably by disruption of HLA-TCR engagement. The effect of mutations within the flanking regions of SARS-CoV-2 epitopes was recently shown by Motozono et al., where they identified a flanking mutation, G446S, in the Omicron BA.1 variant spike protein which increased CD8⁺ T Cell responses [36]. Here, we show that mutations within the nucleoprotein also influence CD8⁺ T cell responses due to differences in the proteasomal processing of viral antigens. Most importantly, we show that mutations such as NP-Q7K are capable of decreasing T cell responses and potentially leading to escape through poor proteasomal processing of the epitope while other mutations are capable of increasing T cell responses. This work highlights the need to study T cell responses in a model system that allows for natural processing and presentation of epitopes.

We established a system which enabled us to evaluate T cell responses to naturally processed epitope peptides and to natural mutations that impact the delivery of epitope antigen to the cell surface even when the epitope sequence is conserved. We, and others [37,38], believe that continual monitoring of potential escape from immunodominant T cell responses in circulating variants before they become dominant strains is critically important for COVID-19 surveillance and future vaccine strategies. Our antigen processing assay was applicable to SARS-CoV-2 encoded NP, but we faced challenges when examining processing of epitopes derived from the Spike protein with very low transduction efficiency despite several optimisation steps. Spike is membrane associated and may be processed via the endocytic-lysosomal pathway, therefore requiring cross-presentation for CD8⁺ antigen recognition [39,40]. A different experimental approach is necessary to assess the effect of Spike mutations on T cell responses.

While it was encouraging to note that these mutations directly affected generation of the 9mer epitope peptides by the proteasome, in agreement with our T cell data, we accept that we have not established whether the 9mer peptide generated by the proteasome is the most antigenic version. Further studies will be required to determine the transport efficiency of the 9mer or extended peptide antigen variants into the endoplasmic reticulum (ER) via TAP, followed by trimming by ERAP1/2 and loading onto HLA Class I molecules. However, the correlation between 9mer generation by the proteasome and T cell responses are remarkable. For NP-P6L, the proportion of 12mer peptide generated is increased considerably. The mutation of the Proline residue to Leucine may make this 12mer peptide a better substrate for ERAP1/2 trimming, thus increasing the epitope peptide pool, especially as ERAP1/2 do not like to cleave after Proline residues. Potentially, the change in proteasomal cleavage due to the mutation is increasing the pool of peptides suitable for ERAP trimming and thus further increasing the generation of epitope peptides.

We showed that both NP-P6L and NP-D103Y could increase CD8⁺ T cell responses due to increased efficiency of proteasomal digestion suggesting that some variants identified in the NP sequence could be beneficial for viral fitness, representing an advantage over increasing CD8⁺ T cell responses. We believe continued identification of such “beneficial” natural mutations within immunodominant T cell epitope flanking regions could provide a new avenue of research for vaccine design, especially as the inclusion of NP into future vaccine designs against SARS-CoV-2 shows promise [41–43]. The fact that our experimental results, particularly for P6L and Q7K, reflected the predicted *in silico* results suggests that analysis similar to that conducted here could be used on a much larger scale with the most up-to-date sequence information to identify and determine other flanking residue mutations that may also increase

T cell responses. Accumulation of these mutations into the next vaccine design could increase T cell immunity and might provide greater protection against the virus.

The number of mutations investigated in this study reflects a small subset due to the early timepoint of sequence sampling, but three out of the twenty-nine mutations (~10%) originally identified showed significant differences in CD8⁺ T cell responses. Moreover, four mutations (~14%) investigated for their effect on proteasomal digestion showed differences in the amount of 9mer epitope peptide generated. This suggests that there are many mutations that may be naturally occurring in the flanking regions of epitopes that have the potential to alter T cell responses. Future studies into the effect of flanking mutations will hopefully create a bank of mutations for multiple HLA Class I allotypes that could be used to investigate whether these mutations have a collective effect on an individual's or population's response to the virus in terms of disease severity or long COVID.

Our results emphasise that the antiviral efficacy and cross-reactive potential of CD8⁺ T cells should be assessed in response to virus-infected cells, or minimally within systems that allow natural processing of epitopes. Tracking and characterisation of epitope-flanking regions in a system where natural processing of viral antigens occurs will allow us to make more precise judgements of the risk of future viral strains. We also provide a framework for future, larger-scale investigations into the potential effect of flanking residue mutations on T cell responses that could help to create a more robust vaccine in the future.

4. Star methods

4.1. Resource availability

Lead Contact: Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts, Tao Dong (tao.dong@ndm.ox.ac.uk) or Dannielle Wellington (dannielle.wellington@rdm.ox.ac.uk).

Data and code availability: The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [44] partner repository with the dataset identifier PXD032054. Any additional information required to re-analyse the data reported in this paper is available from the lead contact upon request.

4.2. Method details

4.2.1. Identification of mutations of interest proximal to SARS-CoV-2 CD8 nucleoprotein epitopes

On July 6, 2020, all available sequences for SARS-CoV-2 that differed from the original Wuhan strain were downloaded from the GISAID database (www.gisaid.org). The regions NP₄₋₂₂ and NP₁₀₀₋₁₁₉ were analysed for mutations. Those falling within the epitope sequence region or found in less than 5 sequences were discounted for further analysis.

4.2.2. Prediction tool analysis of proteasomal digestion

Using the NetChop 20S 3.0 prediction tool (www.jeddb.org) [27], we evaluated the predicted proteasomal cleavage sites for each peptide region in both WT and mutated sequences (NP₄₋₂₂: [Supplementary Table 2](#), NP₁₀₀₋₁₁₉: [Supplementary Table 3](#)). Data was presented as fold-change compared to the predicted proteasomal cleavage in WT sequences ([Fig. 1B–C](#)).

4.3. Tracking of mutation frequencies

The frequency of mutations over time was assessed from data downloaded from the COVID-19 Viral Genome Analysis Pipeline (<https://cov.lanl.gov/content/index>) tracking mutations tool [45]. Data was accessed on August 1, 2022.

4.3.1. Lentivirus construction & production

SARS-CoV-2 nucleoprotein (NP) was cloned from pHRSIN (kind gift from Prof. Alain Townsend) into the Addgene 17,448 lentivirus vector by InFusion PCR (TaKaRa) and verified by Sanger sequencing. NP expression was coupled to eGFP expression using a T2A linker. Single nucleotide mutations were introduced using site-directed mutagenesis (SDM) (New England BioLabs, NEB Base Changer kit). Once confirmed, mutated gene cassettes were extracted by restriction digest and cloned into a clean vector to prevent any off-target effects of SDM-PCR.

Lentivirus was produced using HEK293T cells, expression constructs were co-transfected with psPAX2 and pMD2. G using PEIPro. Lentiviral supernatants were filtered using a 0.45 µm CA filter and overlaid on 20% sucrose before ultracentrifugation at 90,000g for 120 min at 4 °C. Pelleted virus was resuspended in PBS, aliquoted and stored at –80 °C.

4.3.2. Lentivirus transduction

Cells from donor C–COV19-005 were used for investigation as they contained the HLA for both NP₉₋₁₇-HLA-B*27:05 and NP₁₀₅₋₁₁₃-HLA-B*07:02. Full HLA Class I type for this donor is: A*01:01, A*02:01, B*07:02, B*27:05, C*01:02 and C*07:02. PBMCs were pulsed with neat Epstein-Barr virus (EBV) for 3 h. Cyclosporine A (CsA) was then added to reach a final concentration of 1 µg/ml. Cells were kept in 1 µg/ml CsA (diluted in RPMI 1640 medium with 10% foetal calf serum (FCS), 2 mM glutamine and 100 mg/ml pen/strep) for 14–28 days until B cell colonies appear. After that, CsA was removed and B cells were cultured in RPMI 1640 medium with 10% foetal calf serum (FCS), 2 mM glutamine and 100 mg/ml pen/strep.

EBV-transformed B Cells were then spin inoculated with lentivirus to generate stable cell lines. In brief, B cells were plated at 5×10^5 cells per well of a 96-well U-bottom plate and supplemented with 5×10^6 TU/ml lentivirus and 8 mg/ml polybrene in RPMI +1% FBS. The plate was spun at 500g, 37 °C for 1 h prior to overnight incubation at 37 °C. After 24 h, B cells were washed and cultured in RPMI +10% FBS until enough cells were obtained for cell sorting. Cells were then bulk sorted for GFP expression on a BD FACSAria III Cell sorter.

4.4. Generation of short-term SARS-CoV-2 specific T cell lines

Short-term SARS-CoV-2-specific T cell lines were established as previously described [2,26]. Briefly, 3×10^6 to 5×10^6 PBMCs were pulsed as a pellet for 1 h at 37 °C with 10 μ M of peptides containing T cell epitope regions and cultured in R10 (RPMI 1640 medium with 10% foetal calf serum (FCS), 2 mM glutamine and 100 mg/ml pen/strep) at 2×10^6 cells per well in a 24-well Costar plate. IL-2 was added to a final concentration of 100U/mL on day 3 and cultured for a further 10–14 days. NP₉₋₁₇-specific Bulk T cells and NP₁₀₅₋₁₁₃ specific Bulk T cells were generated by sorting HLA-B*27:05 NP₉₋₁₇ and HLA-B*07:02 NP₁₀₅₋₁₁₃ Pentamer⁺ CD8⁺ T cells (Proimmune, UK) from short-term cell lines.

The purity of the sorted population was confirmed at 95% purity. The isolated cells were then clonally expanded and cultured *in vitro*. The sorted cells were cultured with 2 million irradiated cone blood cells (Oxford NHS Blood and Transplant Service) for 14 days in H10 medium (RPMI-1640 supplemented with 10% v/v heat-inactivated human AB serum (National Blood Service), 2 mmol/L L-glutamine, 1% v/v (500 U/mL) penicillin/streptomycin (Sigma-Aldrich)), supplemented with recombinant human IL2 (200 U/mL; PeproTech). The T cells were confirmed for antigen specificity using HLA-B*27:05 NP₉₋₁₇ and HLA-B*07:02 NP₁₀₅₋₁₁₃ Pentamer staining, ran on Attune Nxt flow cytometry (Thermo Fisher) and analysed using FlowJo v.10 (TreeStar Inc.). The validated T cells were then stored in several batches for future assays and used for all *in vitro* functional assays. T-cell cultures were tested for *Mycoplasma* monthly, maintained for 2 months for every assay, passaged once after every thaw, and reauthenticated again using the tetramers method after every thawing and passage [46].

4.4.1. Cell culture of transduced B cell lines & T cells

B cell lines were cultured in RPMI (Lonza) + 10% foetal calf serum (Sigma) + 2 mM L-glutamine (Sigma) at 37 °C, 5% CO₂. Bulk T cells were cultured in RPMI (Gibco) + 10% human serum (Sigma) + 2 mM L-glutamine (Sigma) + Penicillin-Streptomycin (Sigma, 100 units penicillin and 100 μ g streptomycin per mL), supplemented with 100 U/ml IL-2 (Sigma).

4.5. Flow cytometry analysis of transduced B cells

Transduced and sorted B cell lines were stained with Zombie-Violet live dead stain (ThermoFisher) and PE-*anti*-HLA-ABC (BioLegend) and run on an Attune NXT Flow cytometer (ThermoFisher). Data was analysed using FlowJo v10 (BD Biosciences).

4.6. Western blot analysis of transduced B cells

Transduced B cell lines were homogenized using RIPA lysis buffer (Thermo Fisher Scientific). Anti-glyceraldehyde 3-phosphate dehydrogenase (GAPDH) antibody clone 6C5 (Merck Millipore, MAB374) was used as control antibody; anti-SARS-CoV-2 nucleocapsid antibody (2 μ g/ μ l, Sino Biological, 40,143-R001) was used to probe NP expression. Anti-PSBM6 (AbCam), anti-LMP2 (AbCam) and anti-LMP7 (AbCam) were used individually at 1 in 1000 dilution to assess proteasomal subunit expression in the transduced B cells.

Primary antibodies were probed by IRDye 680LT goat anti-mouse (Li-Cor, 926–68020) and IRDye 800LT goat anti-rabbit (Li-Cor, 925–32210), and visualized using the iBright FL1000 (Invitrogen). Images from Western blot experiments were analysed by Fiji software for band density and expressed in GraphPad Prism as a percentage of the GAPDH expression. A one-way ANOVA was performed along with Tukey's or Sidak's multiple comparisons tests to measure statistical differences between cell lines.

4.7. Intracellular cytokine staining of T cells

Bulk CD8⁺ T cells specific for NP₉₋₁₇ or NP₁₀₅₋₁₁₃ epitope were co-cultured with transduced B cell lines for 4 h with BD GolgiPlug, BD GolgiStop and PE-*anti*-CD107a. Dead cells were labelled using Fixable Zombie Violet dye from ThermoFisher. Cells were then washed, fixed with Cytofix/Cytoperm (BD Biosciences), and stained with PE-Cy7-*anti*-IFN γ (eBioscience), APC-*anti*-TNF α (eBioscience) and BV711-*anti*-CD8 (Biolegend). Cells were fixed with BD CellFix and run on an Attune NXT Flow Cytometer (ThermoFisher).

For analysis, live CD8⁺ T cells were gated on and analysed for their expression of CD107a, IFN γ or TNF α using FlowJo V10 (BD Biosciences). Using each bulk T cell line as a control for the other we calculated the percentage response of NP₉₋₁₇ over NP₁₀₅₋₁₁₃, and vice versa. This data was then normalised to a percentage of the response to WT NP. Data was plotted and analysed in GraphPad Prism v9.2.0. Statistical analysis of responses was done by one-way ANOVA and Dunnett's multiple comparisons test.

To assess T cell specificity, WT B cells were pulsed with epitope peptide at 0, 0.0032, 0.016, 0.08, 0.4, 2 or 10 μ g/ml for 1-h prior to co-culture with bulk CD8⁺ T cells specific for NP₉₋₁₇ or NP₁₀₅₋₁₁₃ epitope.

4.7.1. Proteasomal digestion of peptides

Peptides were purchased from GenScript (information available in [Supplementary Table 4](#)) and reconstituted to 20 mg/ml in

DMSO. Recombinant human proteasomes (Human 20S Proteasome Protein, E360-050, bio-technie; Human 20S Immunoproteasome Protein, E370-025, bio-technie) and peptides were combined at a ratio of 50:1 for a period of 15, 60 or 180 min prior to the addition of 0.1 μ l formic acid to terminate the function of the proteasome. Incubations were carried out at 37 °C with rotation to increase proteasome activity. Digested samples were frozen and stored at -80 °C until analysis by mass spectrometry.

4.8. Analysis of *in vitro* proteasome antigen processing products by mass spectrometry

Products generated by *in vitro* proteasome digestion of antigen precursor peptides were analysed by nano-flow liquid chromatography tandem mass spectrometry (nLC-MS/MS) using a Dionex Ultimate 3000 (Thermo) coupled to a timsTOF Pro trapped ion mobility spectrometry (TIMS) time-of-flight mass spectrometer (Bruker) using a 75 μ m \times 150 mm C18 column with 1.6 μ m particles (IonOpticks) at a flow rate of 400 nL/min at 40 °C. The gradient used was a 17-min increase from 2% buffer B to 30% buffer B, followed by an increase to 95% B in 30 s and a 2.5-min wash and 5 min re-equilibration at 2% B (A: 0.1% formic acid in water. B: 0.1% formic acid in acetonitrile).

The timsTOF Pro mass spectrometer was operated in data dependent acquisition - parallel accumulation, serial fragmentation (ddaPASEF) mode. The TIMS accumulation and ramp times were set to 100 ms, and mass spectra were recorded from 100 to 1700 m/z , with a 0.85–1.30 Vs/cm² ion mobility range. Precursors were selected for fragmentation from an area of the full TIMS-MS scan that excludes the majority of ions with a charge state of 1+. Those selected precursors were isolated with an ion mobility dependent collision energy, increasing linearly from 27 to 45 eV over the ion mobility range. Four PASEF MS/MS scans were collected per full TIMS-MS scan using a 3-cycle overlap, giving a duty cycle of 0.53 s. Ions were included in the PASEF MS/MS scan if they met the target intensity threshold of 2000 and were sampled multiple times until a summed target intensity of 10,000 was reached. A dynamic exclusion window of 0.015 m/z by 0.015 Vs/cm² was used, and sampled ions were excluded from reanalysis for 24 s.

Raw MS data files were searched against a database containing the full synthetic antigen peptide precursor sequences using Peaks Studio Version 10 with mass tolerances of 20.0 ppm for precursors and 0.04 Da for fragment ions. Nonspecific cleavage was used and oxidation and deamidation were allowed as variable modifications. Peptide abundance was measured as accumulated ion counts after extraction of the chromatographic peak profiles of the precursor masses using Skyline version 21 and Bruker Compass DataAnalysis Version 5.3. Chromatograms were generated using R software package v4.0.1 and ggplot2 v3.3.2 and ggridges v0.5.2. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [44] partner repository with the dataset identifier PXD032054.

5. quantification and statistical analysis

5.1. Statistical analysis of CD8⁺ T cell responses

Experiments were repeated to an n of 3. All data with statistical analysis was analysed in GraphPad Prism by one-way ANOVA using Dunnett's multiple comparisons test. ANOVA summaries: NP₉₋₁₇ CD107a - F = 33.95, P value < 0.0001; NP₉₋₁₇ IFN γ - F = 17.72, P value < 0.0001; NP₉₋₁₇ TNF α - F = 17.87, P value < 0.0001; NP₁₀₅₋₁₁₃ CD107a - F = 4.095, P value = 0.0492; NP₁₀₅₋₁₁₃ IFN γ - F = 1.323, P value = 0.3328; NP₁₀₅₋₁₁₃ TNF α - F = 4.492, P value = 0.0397. Brown-Forsythe tests: NP₉₋₁₇ CD107a - F (DFn, DFd) = 0.6481 (5,12), P value = 0.6685; NP₉₋₁₇ IFN γ - F (DFn, DFd) = 0.8037 (5,12), P value = 0.5683; NP₉₋₁₇ TNF α - F (DFn, DFd) = 1.382 (5,12), P value = 0.2981; NP₁₀₅₋₁₁₃ CD107a - F (DFn, DFd) = 0.5568 (3,8), P value 0.6581; NP₁₀₅₋₁₁₃ IFN γ - F (DFn, DFd) = 0.5392 (3,8), P value = 0.6686; NP₁₀₅₋₁₁₃ TNF α - F (DFn, DFd) = 0.9408 (3,8), P value = 0.4650.

6. Limitations of the study

This work investigated the effect of a small number of natural mutations in SARS-CoV-2, none of which have yet been present in variants of interest or concern. However, as discussed above we believe this pilot study provides evidence that further investigation of mutations in flanking regions is valid.

Funding

Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS), China (grant number: 2018-I2M-2-002).

Medical Research Council UK programme grant for MRC WIMM Human Immunology Unit.

Additional resources

The following websites were used to access and analyse SARS-CoV-2 sequence data:

<https://gisaid.org/>

<https://cov.lanl.gov/content/index>

<https://covariants.org/>

<https://covariants.org/variants/21K.Omicron>.

The following website was used to predict proteasomal cleavage of sequences:

<https://www.iedb.org/>

Author contribution statement

Danielle Wellington: Conceived and designed the experiments; Performed the experiments; Analysed and interpreted the data; Wrote the paper. Zixi Yin, Zhanru Yu, Raphael Heilig, Simon DavisORCID, Roman FischerORCID, Philip HublitzORCID, Ryan BeveridgeORCID, Danning Dong, Guihai Liu, Yanchun Peng, Xuan Yao: Performed the experiments. Suet Ling Felce and Elie Antoun: Analysed and interpreted the data. Benedikt M Kessler: Conceived and designed the experiments; Performed the experiments; Wrote the paper. Tao Dong: Conceived and designed the experiments; Wrote the paper.

Data availability statement

Data associated with this study has been deposited at The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD032054.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank members of the Dong and Kessler groups for helpful discussions. We thank Prof. Alain Townsend and his group for providing valuable reagents for this work.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2023.e20076>.

References

- [1] S. Rosendahl Huber, et al., T cell responses to viral infections - opportunities for Peptide vaccination, *Front. Immunol.* 5 (2014) 171.
- [2] Y. Peng, et al., Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19, *Nat. Immunol.* 21 (11) (2020) 1336–1345.
- [3] Grifoni, A., et al., SARS-CoV-2 Human T Cell Epitopes: Adaptive Immune Response against COVID-19. *Cell Host & Microbe.*
- [4] J.R. Habel, et al., *Suboptimal SARS-CoV-2-specific CD8+ T cell response associated with the prominent HLA-A*02:01 phenotype*, *Proc. Natl. Acad. Sci. USA* 117 (39) (2020) 24384–24391.
- [5] I. Schulien, et al., Characterization of pre-existing and induced SARS-CoV-2-specific CD8 T cells, *Nat. Med.* 27 (2021) 1–8.
- [6] A.P. Ferretti, 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* 53 (5) (2020) 1095–1107.e3.
- [7] A. Grifoni, et al., Targets of T Cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals, *Cell* 181 (7) (2020) 1489–1501.e15.
- [8] K. Tanaka, The proteasome: overview of structure and functions, *Proc. Jpn. Acad. Ser. B Phys. Biol. Sci.* 85 (1) (2009) 12–36.
- [9] J.S. Blum, P.A. Wearsch, P. Cresswell, Pathways of antigen processing, *Annu. Rev. Immunol.* 31 (1) (2013) 443–473.
- [10] E.J. Sijts, P.M. Kloetzel, The role of the proteasome in the generation of MHC class I ligands and immune responses, *Cell. Mol. Life Sci.* 68 (9) (2011) 1491–1502.
- [11] B.J. Van den Eynde, S. Morel, Differential processing of class-I-restricted epitopes by the standard proteasome and the immunoproteasome, *Curr. Opin. Immunol.* 13 (2) (2001) 147–153.
- [12] N. Pishesha, T.J. Harmand, H.L. Ploegh, A guide to antigen processing and presentation, *Nat. Rev. Immunol.* 22 (12) (2022) 751–764.
- [13] I.A. York, et al., The cytosolic endopeptidase, thimet oligopeptidase, destroys antigenic peptides and limits the extent of MHC class I antigen presentation, *Immunity* 18 (3) (2003) 429–440.
- [14] S.-C. Chang, et al., The ER aminopeptidase, ERAP1, trims precursors to lengths of MHC class I peptides by a “molecular ruler” mechanism, *Proc. Natl. Acad. Sci. USA* 102 (47) (2005) 17107–17112.
- [15] F. Paladini, et al., The multifaceted nature of aminopeptidases ERAP1, ERAP2, and LNPEP: from evolution to disease, *Front. Immunol.* 11 (2020) 1576.
- [16] J. Neeftjes, et al., Towards a systems understanding of MHC class I and MHC class II antigen presentation, *Nat. Rev. Immunol.* 11 (12) (2011) 823–836.
- [17] H. Yu, et al., Conserved sequence preferences contribute to substrate recognition by the proteasome, *J. Biol. Chem.* 291 (28) (2016) 14526–14539.
- [18] R. Abele, R. Tampé, Function of the transport complex TAP in cellular immune recognition, *Biochim. Biophys. Acta Biomembr.* 1461 (2) (1999) 405–419.
- [19] E. Zervoudi, et al., Probing the S1 specificity pocket of the aminopeptidases that generate antigenic peptides, *Biochem. J.* 435 (2) (2011) 411–420.
- [20] T.I. de Silva, et al., The impact of viral mutations on recognition by SARS-CoV-2 specific T cells, *iScience* 24 (11) (2021), 103353.
- [21] M. Del Val, et al., Efficient processing of an antigenic sequence for presentation by MHC class I molecules depends on its neighboring residues in the protein, *Cell* 66 (6) (1991) 1145–1153.
- [22] S.R.F. Ranasinghe, et al., The antiviral efficacy of HIV-specific CD8+ T-cells to a conserved epitope is heavily dependent on the infecting HIV-1 isolate, *PLoS Pathog.* 7 (5) (2011), e1001341.
- [23] R. Draenert, et al., Immune selection for altered antigen processing leads to cytotoxic T lymphocyte escape in chronic HIV-1 infection, *J. Exp. Med.* 199 (7) (2004) 905–915.
- [24] K.M. Peck, A.S. Lauring, Complexities of viral mutation rates, 17, *J. Virol.* 92 (14) (2018), e01031.
- [25] F. Robson, et al., Coronavirus RNA proofreading: molecular basis and therapeutic targeting, *Mol. Cell* 79 (5) (2020) 710–727.

- [26] Y. Peng, et al., An immunodominant NP105–113-B*07:02 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease, *Nat. Immunol.* (2021).
- [27] R. Vita, et al., The immune epitope database (IEDB): 2018 update, *Nucleic Acids Res.* 47 (D1) (2019) D339–d343.
- [28] L. Grenga, et al., Shotgun proteomics analysis of SARS-CoV-2-infected cells and how it can optimize whole viral particle antigen production for vaccines, *Emerg. Microb. Infect.* 9 (1) (2020) 1712–1721.
- [29] Y. Finkel, et al., The coding capacity of SARS-CoV-2, *Nature* 589 (7840) (2021) 125–130.
- [30] Y. Zhao, et al., A dual-role of SARS-CoV-2 nucleocapsid protein in regulating innate immune response, *Signal Transduct. Targeted Ther.* 6 (1) (2021) 331.
- [31] S. Murata, et al., The immunoproteasome and thymoproteasome: functions, evolution and human disease, *Nat. Immunol.* 19 (9) (2018) 923–931.
- [32] M.K. McCarthy, J.B. Weinberg, The immunoproteasome and viral infection: a complex regulator of inflammation, *Front. Microbiol.* 6 (21) (2015).
- [33] P. Cascio, et al., 26S proteasomes and immunoproteasomes produce mainly N-extended versions of an antigenic peptide, *EMBO J.* 20 (10) (2001) 2357–2366.
- [34] P.-M. Kloetzel, The proteasome and MHC class I antigen processing, *Biochim. Biophys. Acta Mol. Cell Res.* 1695 (1) (2004) 225–233.
- [35] J.A. López de Castro, How ERAP1 and ERAP2 shape the peptidomes of disease-associated MHC-I proteins, *Front. Immunol.* 9 (2463) (2018).
- [36] C. Motozono, et al., The SARS-CoV-2 Omicron BA.1 spike G446S mutation potentiates antiviral T-cell recognition, *Nat. Commun.* 13 (1) (2022) 5440.
- [37] Vardhana, S., et al., Understanding T-cell responses to COVID-19 is essential for informing public health strategies. *Science Immunology*.0(0): p. eabo1303..
- [38] D. Wellington, et al., Immunodominance complexity: lessons yet to be learned from dominant T cell responses to SARS-COV-2, *Curr Opin Virol* 50 (2021) 183–191.
- [39] R. Parker, et al., Mapping the SARS-CoV-2 spike glycoprotein-derived peptidome presented by HLA class II on dendritic cells, *Cell Rep.* 35 (8) (2021), 109179.
- [40] C. Qiu, et al., CD8+ T-cell epitope variations suggest a potential antigen HLA-A2 binding deficiency for spike protein of SARS-CoV-2, *Front. Immunol.* 12 (2022).
- [41] T. Dangl, et al., Combining spike- and nucleocapsid-based vaccines improves distal control of SARS-CoV-2, *Cell Rep.* 36 (10) (2021), 109664.
- [42] W.E. Matchett, et al., Cutting edge: nucleocapsid vaccine elicits spike-independent SARS-CoV-2 protective immunity, *J. Immunol.* 207 (2) (2021) 376–379.
- [43] E.K.V.B. Silva, et al., Immunization with SARS-CoV-2 Nucleocapsid protein triggers a pulmonary immune response in rats, *PLoS One* 17 (5) (2022), e0268434.
- [44] Y. Perez-Riverol, et al., The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences, *Nucleic Acids Res.* 50 (D1) (2022) D543–d552.
- [45] B. Korber, et al., Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus, *Cell* 182 (4) (2020) 812–827.e19.
- [46] M. Abd Hamid, et al., Self-maintaining CD103+ cancer-specific T cells are highly energetic with rapid cytotoxic and effector responses, *Cancer Immunol. Res.* 8 (2) (2020) 203–216.