

# Mapping the immunopeptidome of seven SARS-CoV-2 antigens across common HLA haplotypes

Received: 5 November 2023

Accepted: 21 August 2024

Published online: 30 August 2024

 Check for updates

Asolina Braun<sup>1</sup>, Louise C. Rowntree<sup>2</sup>, Ziyi Huang<sup>1</sup>, Kirti Pandey<sup>1</sup>, Nikolas Thuesen<sup>3</sup>, Chen Li<sup>1</sup>, Jan Petersen<sup>1</sup>, Dene R. Littler<sup>1</sup>, Shabana Raji<sup>2</sup>, Thi H. O. Nguyen<sup>2</sup>, Emma Jappe Lange<sup>3</sup>, Gry Persson<sup>3</sup>, Michael Schantz Klausen<sup>3</sup>, Jens Kringelum<sup>3</sup>, Shanzou Chung<sup>1</sup>, Nathan P. Croft<sup>1</sup>, Pouya Faridi<sup>1</sup>, Rochelle Ayala<sup>1</sup>, Jamie Rossjohn<sup>1,4</sup>, Patricia T. Illing<sup>1</sup>, Katherine E. Scull<sup>1</sup>, Sri Ramarathinam<sup>1</sup>, Nicole A. Mifsud<sup>1</sup>, Katherine Kedzierska<sup>2</sup>, Anders Bundgård Sørensen<sup>3</sup> & Anthony W. Purcell<sup>1</sup>✉

Most COVID-19 vaccines elicit immunity against the SARS-CoV-2 Spike protein. However, Spike protein mutations in emerging strains and immune evasion by the SARS-CoV-2 virus demonstrates the need to develop more broadly targeting vaccines. To facilitate this, we use mass spectrometry to identify immunopeptides derived from seven relatively conserved structural and non-structural SARS-CoV-2 proteins (N, E, Nsp1/4/5/8/9). We use two different B-lymphoblastoid cell lines to map Human Leukocyte Antigen (HLA) class I and class II immunopeptidomes covering some of the prevalent HLA types across the global human population. We employ DNA plasmid transfection and direct antigen delivery approaches to sample different antigens and find 248 unique HLA class I and HLA class II bound peptides with 71 derived from N, 12 from E, 28 from Nsp1, 19 from Nsp4, 73 from Nsp8 and 45 peptides derived from Nsp9. Over half of the viral peptides are unpublished. T cell reactivity tested against 56 of the detected peptides shows CD8<sup>+</sup> and CD4<sup>+</sup> T cell responses against several peptides from the N, E, and Nsp9 proteins. Results from this study will aid the development of next-generation COVID vaccines targeting epitopes from across a number of SARS-CoV-2 proteins.

The initial vaccines designed to combat COVID-19 were predominantly focussed on eliciting strong antibody-mediated immune responses against the original Wuhan strain Spike protein. Whilst this focussed approach led to the generation of several highly effective, safe vaccines within an astonishingly brief timeframe, it also comes with some limitations. For one, the SARS-CoV-2 virus has acquired various mutations, some of which result in an altered Spike protein which

impacts vaccine efficacy<sup>1,2</sup>. Additionally, it is becoming increasingly clear that apart from eliciting strong antibody responses, successful recruitment of T cells, especially those recognising conserved viral epitopes, is highly beneficial due to their strain cross-reactivity<sup>3-5</sup>. Conversely, the adverse health effects of long COVID are associated with dysregulated T cell and humoral responses<sup>6</sup>. Efficient anti-SARS-CoV-2 T cell responses can be generated upon natural infection or

<sup>1</sup>Department of Biochemistry and Molecular Biology, Infection and Immunity Program, Biomedicine Discovery Institute, Monash University, Clayton, VIC, Australia. <sup>2</sup>Department of Microbiology and Immunology, University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Melbourne, VIC, Australia. <sup>3</sup>Evaxion Biotech, Hørsholm, Denmark. <sup>4</sup>Institute of Infection and Immunity, Cardiff University, School of Medicine, Cardiff, UK.

✉ e-mail: [anthony.purcell@monash.edu](mailto:anthony.purcell@monash.edu)

vaccination with improved responses after booster vaccination<sup>7</sup>. However, currently T cell immunity elicited after vaccination wanes within 6–12 months while the half-life of T cell responses after natural infection is over one year<sup>7,8</sup>. This indicates that an improvement of vaccine design to purposefully incorporate and drive a stronger T cell defence would be advantageous.

The first step of generating adaptive immune responses against pathogens, is the processing of foreign proteins into short fragments (peptide antigens), which are presented to immune cells by Human Leukocyte Antigen (HLA) class I and class II molecules. Strong antibody responses are the prime target of vaccines and rely on antigen-specific HLA class II-mediated presentation to CD4<sup>+</sup> T cells, which subsequently activate B cells thus leading to antibody generation. Whilst antibody levels against SARS-CoV-2 diminish within the first few months post-vaccination, T cell immunity remains functional for over six months following infection<sup>7,9,10</sup>. It is now also understood that emerging SARS-CoV-2 variants are, to a degree, able to escape humoral immunity, but can still be recognised by the T cell arm of the immune system<sup>11</sup>. Whilst antibody-mediated immune responses rely on HLA class II antigen presentation to CD4<sup>+</sup> T cells, cytotoxic CD8<sup>+</sup> T cell responses are primed via the HLA class I antigen presentation pathway and can provide potent protection against infection by eradicating infected cells and preventing viral spreading<sup>12–15</sup>. The identification and study of HLA-bound peptides (immunopeptidomics) is possible using specialised, mass spectrometry centred protocols<sup>16</sup>. Yet, immunopeptidomic data of SARS-CoV-2 epitopes is still relatively scarce. Currently, the field mostly relies on the prediction of HLA class I and class II SARS-CoV-2 peptide epitopes and directed peptide mapping of T cell reactive epitopes to instruct vaccine design<sup>17–20</sup>. We have previously developed the RAVEN<sup>TM</sup> AI model to design T cell epitope-based viral vaccines and shown that a RAVEN<sup>TM</sup> designed vaccine protects against severe symptoms of SARS-CoV-2 infection in mice<sup>13</sup>. However, the prediction of natural processing and presentation of viral peptides by RAVEN<sup>TM</sup> and other tools requires functional or immunopeptidomic-based training and validation with few ground truth datasets available to refine algorithms for SARS-CoV-2<sup>20,21</sup>. Immunopeptidomics is a mass spectrometry-based approach that can uncover experimentally derived peptide antigens presented by HLA molecules that have successfully undergone various stages of antigen processing such as proteasome and aminopeptidase cleavage, endoplasmic reticulum trafficking and HLA binding<sup>16</sup>.

Overall, there are only a few SARS-CoV-2 immunopeptidomic studies, two of which focus solely on the Spike protein and its presentation by HLA class II molecules<sup>22,23</sup>. The Sabeti group investigated the HLA class I immunopeptidome of SARS-CoV-2 infected HEK293T and A549 cell lines, with a total of 25 unique SARS-CoV-2 peptides reported<sup>24</sup>. More recently the same group has investigated the HLA-DR class II immunopeptidome of infected A549 and HEK293 cells showing a strong bias towards the presentation of viral structural proteins (M, N, S)<sup>25</sup>. The Samuels group reported 26 HLA class I- and 36 HLA class II-restricted SARS-CoV-2 peptides using infected cells (IHW01070, HEK293T, and Calu-3) and overexpression of SARS-CoV-2 genes in 721.221 B-lymphoblastoid cells with monoallelic HLA expression<sup>26</sup>. A further report from the Yee group identified five epitopes from the non-structural protein (Nsp) 13 and the membrane protein<sup>27</sup>. In contrast, rather than focussing on SARS-CoV-2 derived peptides, Yin and colleagues describe COVID-induced changes in the immunopeptidome that reflect the immunopathology in lung biopsies following lethal infection<sup>28</sup>. A hallmark of these studies is the relative paucity of virus-derived peptides compared to the self-repertoire and limited HLA allotype coverage due to restricted cell line use or limited antigen coverage.

Here we present 200 SARS-CoV-2-derived unique native peptides and post-translational modifications thereof. Notably, these peptides are presented by HLA molecules widely represented in the global

population, including class I HLA-A\*01:01 (4.8%) and A\*02:01 (15.3%) and class II HLA-DQB1\*02:01 (15.0%) and DPBI\*04:01 (23.3%)<sup>29</sup>. Our work characterises peptides from the nucleocapsid, envelope protein, Nsp1 (host translation inhibitor), Nsp4 (transmembrane protein, part of the viral replication-transcription complex), Nsp5 (main protease), Nsp8 (co-factor of the RNA dependent RNA polymerase) and Nsp9 (part of the viral replication-transcription complex). Overall, our findings uncover a significant number of naturally processed SARS-CoV-2 immunopeptides which can inform the design of second-generation vaccines targeting humoral as well as cellular immunity against a mix of SARS-CoV-2 antigens.

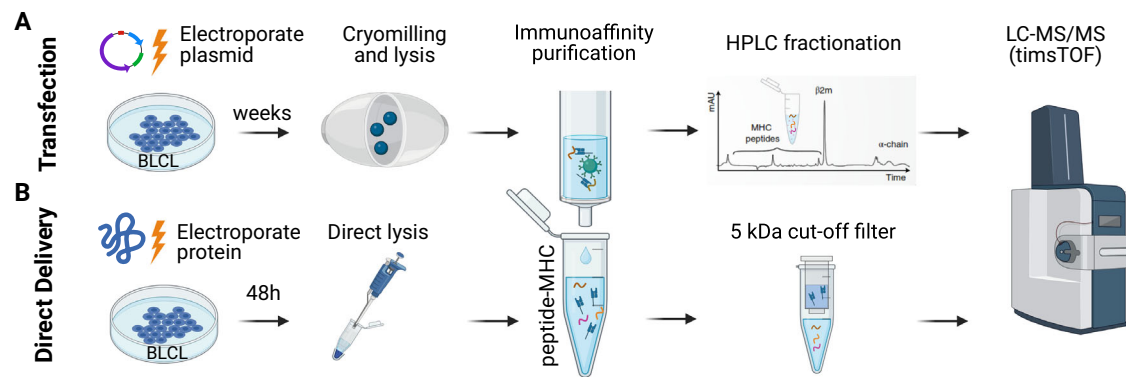
## Results

### Applying complementary approaches to uncover SARS-CoV-2 epitopes

To explore the SARS-CoV-2 derived immunopeptidome presented by endogenous HLA molecules in human cells we chose B lymphoblastoid cell lines (BLCL) that are known to express high levels of class I and class II HLA in the context of naturally occurring HLA haplotypes (Supplementary Fig. S1A). B cells are biologically relevant as they play a central role in the adaptive immune response and the specific cell lines chosen carry HLA allotypes that are highly representative of substantial proportions of the human population (Supplementary Fig. S1B). In this study we have explored the conventional transfection approach of cells with plasmids encoding SARS-CoV-2 *nucleocapsid (N)*, *envelope (E)*, *Nsp8*, and *Nsp9* genes. Together with the structural N and E proteins, Nsp8 and Nsp9 are among the most highly abundant non-structural proteins detectable from a few hours post infection with continuing expression until at least 24 h post infection<sup>30</sup>. The transfection approach with plasmids containing a fluorescent marker allows the successfully transfected cells to be sorted, facilitating the establishment of stable cell lines of high purity. Such cells can be readily expanded in culture allowing the collection of a high number of cells conducive to deep coverage immunopeptidomics datasets that facilitate the discovery of rare peptide subsets, such as from SARS-CoV-2. However, this approach requires approximately two months and substantial tissue culture resources to complete. Hence, we also explored an alternative approach of directly delivering SARS-CoV-2 proteins into the cell. We hypothesised this approach would be capable of directing the viral proteins into both, class I and class II antigen presentation pathways, thereby maximising the detection of not only HLA class II, but also HLA class I derived SARS-CoV-2 epitopes. Electroporation of purified recombinant viral and other proteins is a time-efficient, established technique with the steps from protein delivery into mammalian cells to cell harvest only requiring 48 h. We reasoned that if such a delivery method could make the electroporated protein accessible to both antigen processing pathways, it would be of great use in future antigen discovery settings. Such an expedited approach would facilitate timely epitope discovery and vaccine development when novel pathogens or pathogen variants are encountered. The individual immunopeptidomics workflows for BLCL undergoing direct delivery of antigen and BLCL stable transfectants are depicted in Fig. 1 and follow previously described methodologies<sup>16,31</sup>.

### Characterisation of HLA ligands presented by BLCL transfected with SARS-CoV-2 derived genes

We used mass-spectrometry-based immunopeptidomic analysis to identify epitopes of the viral E, N, Nsp8, and Nsp9 proteins in stably transfected 9004 and 9087 BLCL cell lines. In all cases, raw data was searched using PEAKS 10 online against a human and SARS-CoV-2 combined proteome database. We detected an average of 34,164 8- to 12-mer human peptides isolated using the anti-pan class I antibody W6/32 and an average of 76,863 10- to 20-mer peptides per sample isolated using a mix of HLA class II specific antibodies (SPV-L3, B7/21 and LB3.1) with 5% FDR cut-off applied (Fig. 2A). In total, in excess of 500,000



**Fig. 1 | Workflow for the generation of SARS-CoV-2-derived HLA I and HLA II immunopeptidome data.** **A** Upper panel depicts the large-scale workflow used for the stable transfection and processing of B lymphoblastoid cell lines (BLCL) cells using EF1 $\alpha$ -Gol-pIRES-DsRed plasmids containing SARS-CoV-2 genes. Cell pellets ( $4 \times 10^8$  to  $1 \times 10^9$ ) were homogenised using a cryomill, lysed, and peptide-HLA complexes immunoaffinity purified using first the W6/32 antibody (HLA class I) and subsequently a mix of SPV-L3, LB3.1, B7/21 antibodies (HLA class II). Peptides were separated from HLA using reverse-phase high-performance liquid chromatography (HPLC). **B** The lower panel shows the workflow schematic for small-scale protein

antigen direct delivery experiments. SARS-CoV-2-derived proteins were electroporated into BLCL and  $0.9 \times 10^8$  to  $1.6 \times 10^8$  cells collected 48 h later. Cell pellets were subjected to direct lysis and immunoaffinity purification. Peptides were separated from HLA using 5 kilodalton (kDa) molecular weight cut-off filters. All samples were analysed using liquid chromatography-tandem mass spectrometry (LC-MS/MS) and PASEF<sup>®</sup> on TimsTOF Pro. MHC: Major histocompatibility complex. Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license.

nonredundant mammalian peptide sequences were identified. Collectively, this deep coverage of the BLCL immunopeptidome translated into the detection of 181 non-redundant SARS-CoV-2 derived peptides (Supplementary Data S1). The non-structural proteins Nsp8 and Nsp9, which are not currently in focus for vaccine development, are efficiently presented: 78% of the Nsp8 and 85% of the Nsp9 sequences were covered by the 9004 and 9087 BLCL immunopeptidomes.

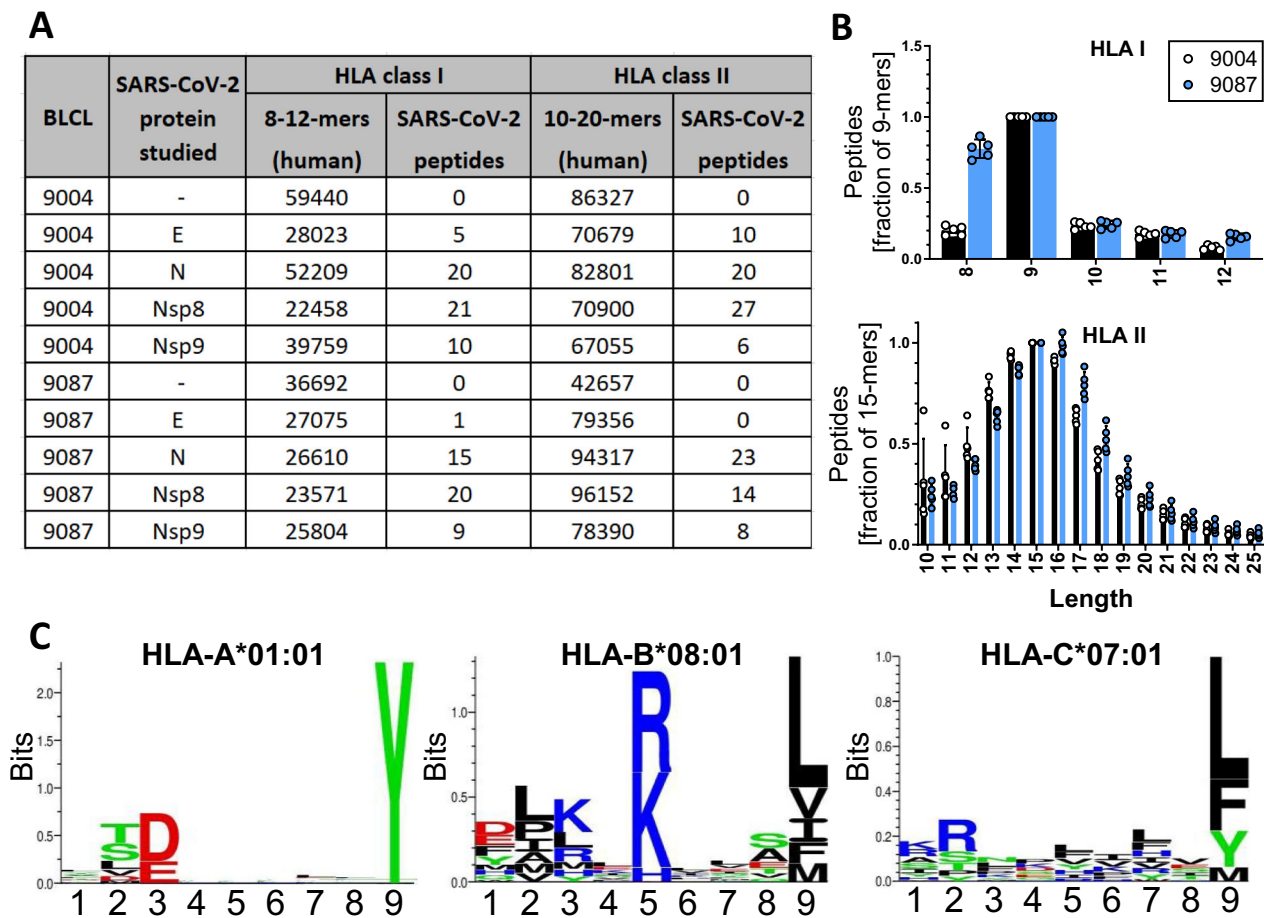
The overall length distribution of peptides eluted from class I and class II HLA molecules are shown in Fig. 2B. As expected, peptides presented by HLA class I were predominantly 9 amino acids in length, while HLA class II bound peptides were mainly 12–18 amino acids long. It is noteworthy that for the 9087 BLCL we additionally detected a comparatively high number of 8-mer peptides in the HLA class I dataset. Closer analysis revealed that the majority of 8-mer peptides in 9087 were attributable to the HLA-B\*08:01 allotype (Supplementary Fig. S2A, B). Although HLA-B\*08:01 is known to bind to 8-mer peptides<sup>32</sup>, our result was somewhat unexpected as previous studies have reported the number of HLA-B\*08:01 derived 8-mers in proportion to 9-mers to constitute only up to a factor of 0.4 of peptides versus a factor of 0.8 in our dataset<sup>33,34</sup>. Notably, the trapped ion mobility cell of the Bruker timsTOF Pro instrument allowed ready MS/MS interrogation of singly charged precursors by targeting defined regions of the mobiligram (m/z vs ion mobility). Targeting singly charged precursors for MS/MS is particularly important for the detection of shorter non-tryptic peptides. We therefore analysed the charge state of 8-mers in our dataset and indeed confirmed that the majority carried a single charge in contrast to 9-mers which typically acquired a double charge (Supplementary Fig. S2C). These singly charged peptides likely remained undetected when using alternative mass spectrometry approaches in earlier studies. Next, we performed binding motif analysis, the main anchor residues were in line with the expected binding motifs of HLA-I allotypes expressed in the cell lines, and a representative 9087 elution is shown (Fig. 2C). In summary, this dataset not only serves its primary purpose of detecting a high number of previously unknown SARS-CoV-2 derived peptides, but also represents a valuable resource as an extended human immunopeptidome study of BLCLs with an expanded peptide detection for particularly 8-mers of the HLA-B\*08:01 allotype (Supplementary Data S2, S3; currently documented here: <https://virusms.erc.monash.edu/browse.jsp>).

### HLA-peptide identification using direct delivery of SARS-CoV-2-derived proteins

As an alternative approach, we chose to deliver SARS-CoV-2 proteins via electroporation into BLCL cells (direct delivery). Nsp9 and N proteins were used as representatives of shorter and longer proteins to compare both approaches. Additionally, Nsp1, Nsp4, and Nsp5 were delivered into BLCL to expand the number of investigated SARS-CoV-2 proteins in this study. On average, 10,578 8- to 12-mers were identified in HLA class I eluted samples and 22,044 10- to 20-mers in HLA class II eluted samples (Fig. 3A). These numbers are in line with the lower cell input used for direct delivery experiments compared to the transfected cells. The length distribution of eluted peptides was very similar to the previously detected length preference of transfected cells (Fig. 3B). The peptide binding analysis for the direct delivery approach was a close match to binding motifs from transfectants, a representative 9004 elution is shown (Fig. 3C). Of note, in 9004 BLCL the HLA-A\*02:01 motif was predominant so that a separation of motifs using Gibbs clustering was not applicable. Overall, we detected 81 unique SARS-CoV-2 derived peptides from the direct antigen delivery approach stemming from HLA class I and class II processing. None of the viral peptides from the transfection nor direct delivery approach were detected in parental cell lines that were used as negative controls.

### Comparison of direct delivery and stable transfection approaches

We next aimed to further examine the data for SARS-CoV-2 N protein and Nsp9 for which we have acquired datasets from both, direct delivery and stable transfection approaches. For N protein, we detected a total of 32 and 33 peptides from HLA class I and II processing pathways in stably transfected 9004 and 9087 BLCL respectively. The outcome was markedly different for direct delivery where we found no N-derived peptides in 9004 BLCL and ten peptides exclusively in the HLA class II elution of 9087 BLCL. The observed differences may be caused by a preference for direct loading of HLA class I molecules for nucleocapsid-derived peptides, with processivity issues of the antigen resulting in poor presentation with direct antigen delivery which may proceed predominantly through the HLA class II endolysosomal pathway. The presentation of Nsp9-derived ligands was more comparable between both methods. Peptide matches featuring the same 9-mer core were found by both direct antigen delivery and stable



**Fig. 2 | Immunopeptidomic profile of class I and class II eluted peptides presented by stably transfected BLCL.** Samples were analysed using PEAKS Online using a 5% FDR cut-off. **A** Relevant eluted peptide numbers, headers in bold. **B** Mean length distribution of class I and class II eluted peptides, SD is shown, based on all  $n = 5$  cell lines in A; 9087: black bars, 9004: blue bars. **C** Representative Gibbs

Cluster peptide motif analysis of 9-mers from the HLA class I elution of 9087\_E is shown. Attribution of individual clusters to known HLA class I alleles expressed in 9087 is shown above each cluster, clusters are based on 1130-5414 peptides. Analysis using Immunolyser<sup>53</sup>. Nsp: non-structural protein, BLCL: B lymphoblastoid cell line, HLA: human leukocyte antigen. Source data are provided as a Source Data file.

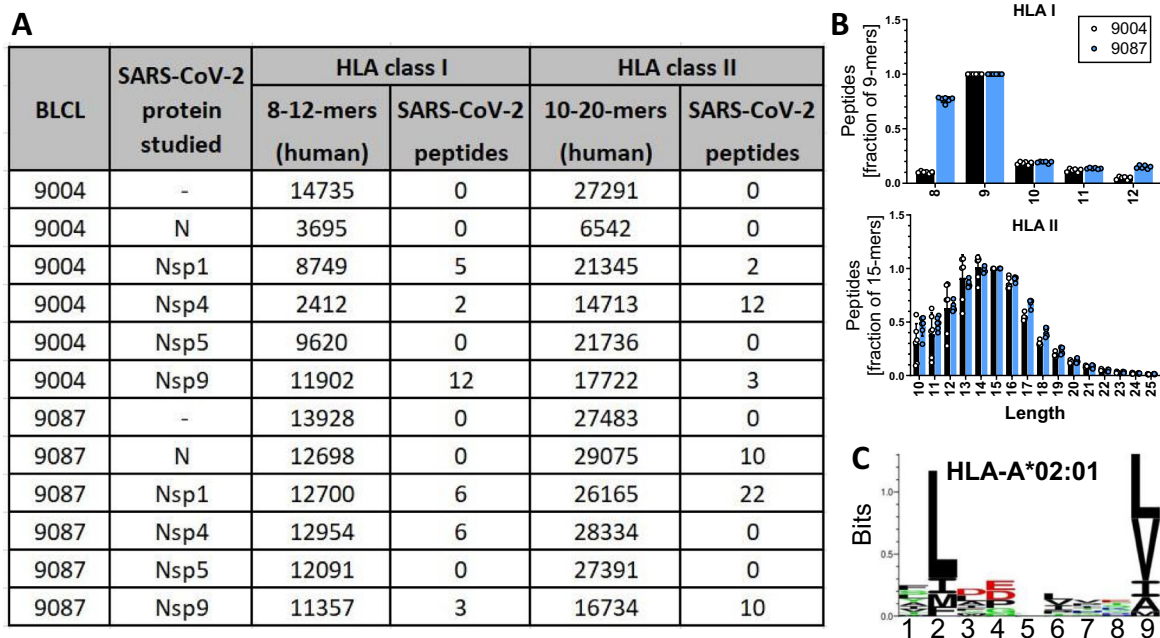
transfected cells with an additional number of peptides detected only in stably transfected cells and some HLA class II peptides detected only via the direct delivery approach (Supplementary Data S1, Fig. 4A). Thus, there was a substantial overlap in the HLA class I- and class II-derived antigens from Nsp9 using direct antigen delivery and stable transfection whereas, for nucleoprotein-derived peptides, the overlap between the methods is restricted to HLA class II-derived peptides. Overall, the exogenous introduction of protein in the direct delivery approach directs antigen presentation predominantly to the HLA class II pathway, while the transfection approach allows to detect HLA class I and class II peptides in a more balanced ratio, at least when B cells are studied which express high levels of HLA class II and may be less efficient than other antigen-presenting cells to cross-present antigen. Where time is critical and HLA class II presentation is under investigation, the direct delivery approach provides a clear time advantage but if protein expressed in bacteria is used, this approach might not capture mammalian post-translational modifications. The transfection approach, though slower and more labour-intensive, provides a more complete gamut of HLA class I (and class II) antigens derived from the transfected gene of interest and is critical for approaches where the depth of antigen discovery is key.

#### Immunopeptidomic profiling of viral antigen presentation

The detected viral HLA ligands were aligned to their SARS-CoV-2 source protein to generate an antigen presentation density map

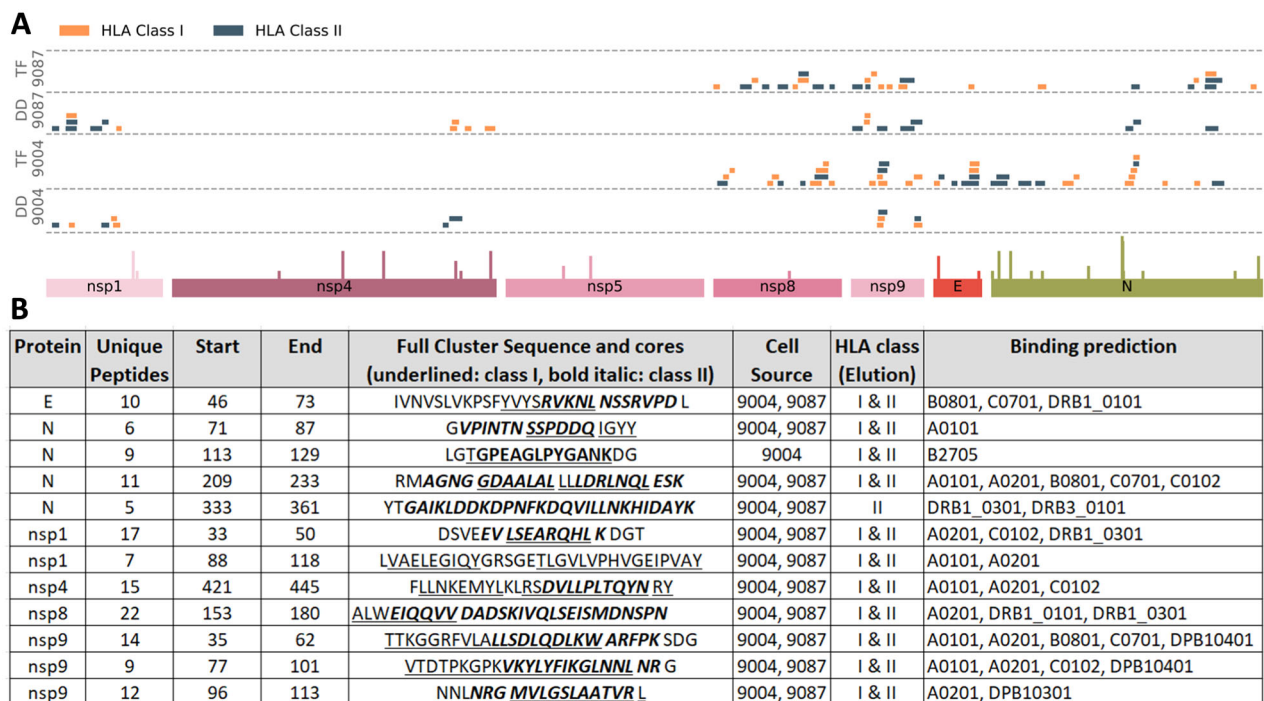
(Fig. 4A). Additionally, mutations found in these antigens in variants of concern are indicated as spikes projecting from the protein bar, thereby rapidly highlighting more conserved epitopes of potentially high interest for vaccine design. The alignment revealed twelve hotspots of antigen presentation with clusters of closely located immunopeptides which are repeatedly detected across the different HLA classes, allotypes and cell lines (Fig. 4B). Collectively, binding prediction to source cell line-specific HLA allotypes confirmed that 53% (105/200) of the detected, PTM-stripped viral immunopeptides were predicted binders to at least one HLA allotype expressed by the cell lines. Specifically, the predicted binding of the eluted peptides to the expressed HLA class I and class II allotypes using netMHCpan and netMHCIIpan assigned 68% of 9-mers and 40% of 15-mers (9004 BLCL) as well as 77% of 9-mers and 83% of 15-mers (9087 BLCL) to at least one of the expressed HLA class I or class II allotypes respectively (Supplementary Data S1).

Due diligence in vaccine design also aims to exclude, as much as possible, any potential cross-reactivity of the pathogenic antigen to human proteins to avoid adverse events such as the development of autoimmunity. We have therefore analysed the detected SARS-CoV-2-derived HLA ligands for similarity to the human genome. Each peptide was allowed up to two mismatched amino acids without insertions or deletions. Using this approach, 43 out of 200 peptides displayed similarity to human protein-derived peptides (Supplementary Data S4).



**Fig. 3 | Immunopeptidomic profile of class I and class II eluted peptides presented by BLCL 48 h after direct delivery of protein antigen.** Samples were analysed using PEAKS Online using a 5% FDR cut-off. **A** Relevant eluted peptide numbers, headers in bold. **B** Mean length distribution of HLA I and HLA II eluted peptides. SD is shown, based on all  $n = 5$  cell lines in A; 9087: black bars, 9004: blue bars. **C** Representative Gibbs Cluster peptide motif analysis of 9-mers from an HLA

class I elution of 9004 with delivery of Nucleoprotein is shown. HLA-A\*02:01-derived peptides dominate the HLA class I immunopeptidome, cluster size is 2602 peptides. Analysis using Immunolyser<sup>53</sup>. Nsp non-structural protein, BLCL B lymphoblastoid cell line, HLA human leukocyte antigen. Source data are provided as a Source Data file.



**Fig. 4 | Distribution of HLA-I and HLA-II presented protein regions within SARS-CoV-2 derived investigated proteins.** **A** SARS-CoV-2 derived peptides found across stable transfected (TF) B lymphoblastoid cell lines (BLCL) and direct delivery (DD) antigen approaches and their alignment to source proteins. Peaks represent

mutations in variants of concern. **B** Table of antigen presentation hotspots. Peptide sequences derived from human leukocyte antigen (HLA) class I elutions are underlined, peptide sequences derived from HLA class II elutions are shown in bold italic. Nsp: non-structural protein, Source data are provided as a Source Data file.

## Post-translational modification (PTM) of peptide ligands

Among all 248 identified non-redundant viral peptides, 54 peptides contained modified residues. Among the detected PTMs were oxidation (M), deamidation (Q and N), dehydration, acetylation, cysteinyl-ation, and others, consistent with known modifications of HLA-presented antigens, in source fragmentation of peptide ions or artefacts induced by sample processing (Supplementary Fig. S3A)<sup>35–38</sup>. Two deamidated Asn-containing peptides are of particular interest, <sup>32</sup>YYN(+0.98)TTKGGRF<sup>41</sup> and <sup>267</sup>AYN(+0.98)VTQAF<sup>276</sup> originating from Nsp9 and nucleoprotein respectively. Both were derived from HLA class I eluted ligands of stably transfected 9004 BLCLs and feature the NX(S/T) motif. We have recently shown that this modification is the result of N-glycan removal and as such is a feature of formerly glycosylated precursors<sup>36</sup>. Indeed, glycoproteomic analysis has previously found that N269 of the nucleocapsid protein has a 94% glycan occupancy rate with the glycan mostly comprised of -85% high-mannose type glycosylation<sup>39</sup>.

## Correlation with reported ligands in the Immune Epitope Database (IEDB)

All reported ligands and epitopes in the IEDB were extracted for SARS-CoV2 and manually curated to produce Fig. 5. A large amount of data was derived from direct/competitive fluorescence binding assays ( $n = 6211$ ), high-throughput multiplex assays ( $n = 5703$ ) and only 378 from MS-based ligandomics compared to the 248 reported here. As can be seen in Fig. 5 where only comparison with MS-based ligandomics is included (see Supplementary Fig. S4 for all assays), we increased the number of data points for HLA-B\*08:01/B\*27:05, HLA-C\*01:02/C\*07:01, HLA-DQA\*05:01/DQB1\*02:01 and HLA-DRA1\*01/HLA-DRB1\*01:01/DRB1\*03:01 to a large degree and to some extent for the two highly prevalent HLA allotypes A\*01:01/A\*01:02. In general, our dataset adds significant data for almost all evaluated proteins in regards to HLA class I and class II allotypes, with N-protein being an exception where significant data was already reported for HLA-A\*01:01/A\*01:02 and HLA-B\*08:01/B\*27:05.

## Immunogenicity of a subset of detected SARS-CoV-2 peptides

In order to further assess the clinical relevance of detected viral peptides, we then proceeded to test them on samples from COVID convalescent individuals. From all detected viral peptides, we selected 56 SARS-CoV-2 peptides based on their predicted HLA binding while excluding peptides containing reactive cysteines. Peptides originating from both, the transfection and direct delivery approaches were synthesised and validated via LC-MS/MS and peptide spectrum matching (Supplementary Fig. S3B–E). These peptides are currently documented in our virusMS database ([https://virusms.erc.monash.edu/browse.jsp;EXP\\_003](https://virusms.erc.monash.edu/browse.jsp;EXP_003));<sup>40</sup>. The synthetic peptides were subsequently distributed into three pools of 17–20 peptides, with nested peptide sets pooled together where possible, and used in T cell activation assays using PBMC collected at 12–223 days post SARS-CoV-2 infection from 9 individuals expressing HLA allotypes that matched the BLCLs (Supplementary Data S5). All donor PBMCs were stimulated with three individual peptide pools and expanded for 10 days before the assessment of SARS-CoV-2-reactive T cells by intracellular cytokine production. Responses against a number of peptide pools were detected across individuals (Fig. 6A, gating in Supplementary Fig. S5). Cultures responding to a pool on day 10 were restimulated on day 13 with individual peptides or nested subsets from peptide pools (Fig. 6B–D). Remarkably, in one donor with PBMC collection at 72 days post-symptom onset, we detected CD4<sup>+</sup> T cell responses against nested peptide pairs derived from E, Nsp1- and Nsp9-proteins (Fig. 6B). The nested peptides E<sub>57-65</sub> (YVYSRVKNL) and E<sub>58-73</sub> (VYSRVKNLNSSRPDL) are both predicted to bind to HLA-DRB1\*01:01, which is expressed by both the source BLCL and the donor. The nested peptides Nsp1<sub>88-99</sub> (LVAELEGIQYGR) and Nsp1<sub>89-97</sub> (VAELEGIQY) could be derived from either HLA-

DRB1\*01:01 or HLA-DQB1\*05:01, which are both shared between the BLCL source (DQAI\*01:01) and donor (DQAI\*01:01/05, DQAI\*03:02/03/09). A further shared allotype, HLA-DPBI\*04:01, is predicted to bind the remaining nested peptides Nsp9<sub>45-55</sub> (LLSDLQDLKWA) and Nsp9<sub>45-59</sub> (LLSDLQDLKWARFPK). The donor and BLCLs also share HLA-DPAI\*01:03 for which both are homozygous.

Strikingly, we could further detect a dominant response across three individuals against the 19-mer peptide N<sub>343-361</sub> (DPNFKDQVILLNKHIDAYK) whilst the overlapping N<sub>333-349</sub> peptide (YTGAIKLDDKDPNFKDQ) did not have a noteworthy contribution to the antigen-specific response against pool 2 (Fig. 6C). PBMC from these three donors 1, 8 and 9 were collected at 109, 210 and 12 days post-symptom onset, respectively. The peptide N<sub>343-361</sub> was detected in a HLA-class II elution from N protein-transfected 9004 BLCL (DRB1\*01:01, DQB1\*05:01, DPBI\*04:01). Donor 1 shares DQB1\*05:01 while donors 8 and 9 share the highly prevalent DPBI\*04:01 allotype. Interestingly, whilst donors 1 and 8 were unvaccinated, donor 9 was a breakthrough infection indicating the initiation of a non-Spike N-specific T cell response post-vaccination. These results indicate that convalescent COVID-19 patients have persisting SARS-CoV-2-specific CD4<sup>+</sup> T cell immunity against N protein for over 7 months post-infection.

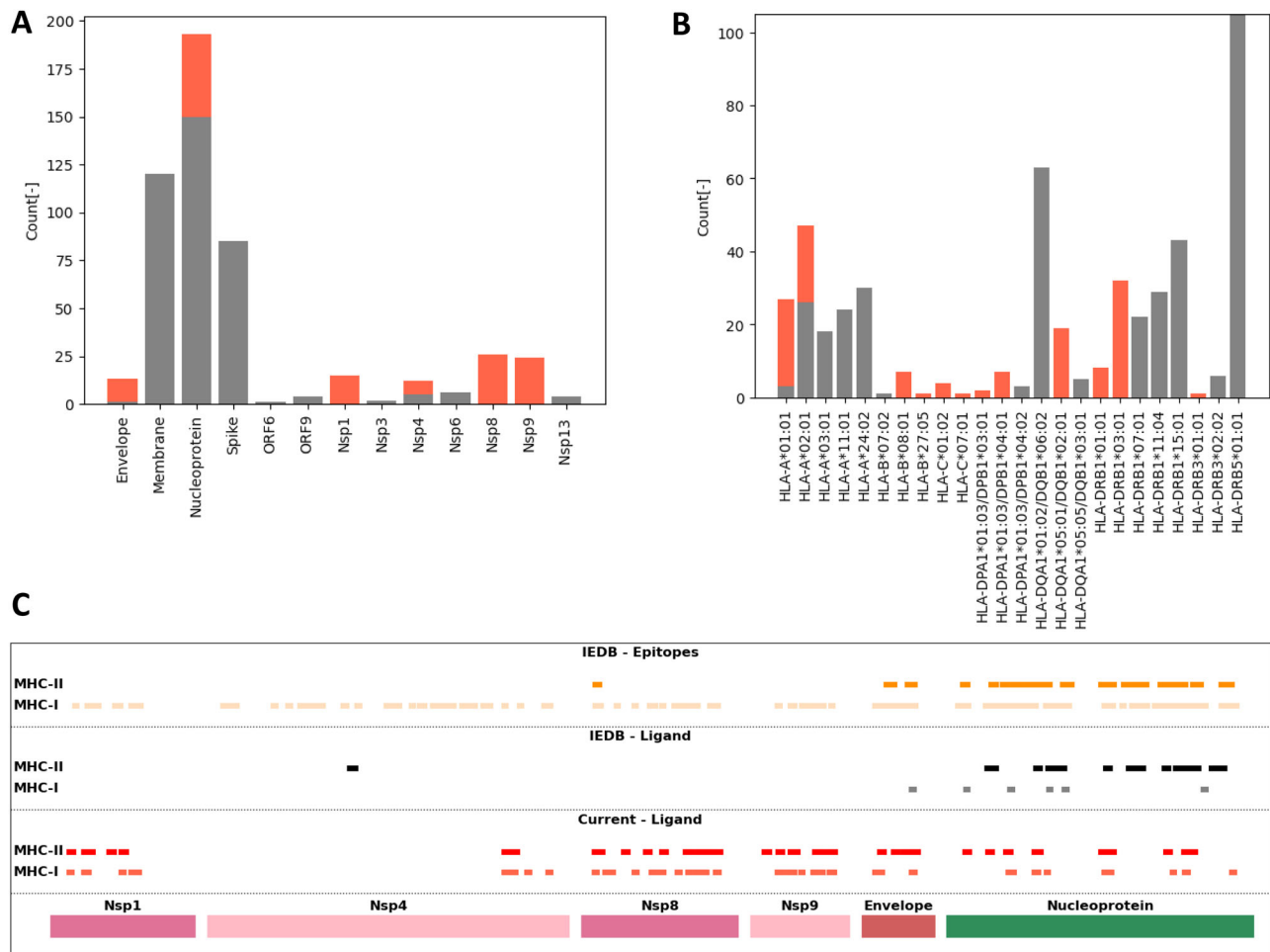
In addition to validating a number of peptides that elicit a CD4<sup>+</sup> T cell response, we also detected a CD8<sup>+</sup> T cell response against pool 3 in one donor where PBMC collection was at 214 days post-symptom onset (Fig. 6D). Due to limited sample size, further tests were carried out on sub-pools. We could narrow down the response to nested sub-pool 3B containing 10–15-mer peptides N<sub>71-85</sub> (GVPINTNPPDDQIG), N<sub>77-86</sub> (NSSPDDQIGY) and N<sub>77-87</sub> (NSSPDDQIGYY). N<sub>77-86</sub> and N<sub>77-87</sub> are both predicted to bind HLA-A\*01:01, an allotype shared between the source BLCL and the immunoreactive PBMC donor.

## Discussion

Over recent years SARS-CoV-2 has become endemic and the disease still continues to pose a high burden on health systems worldwide. This continued burden is mainly caused by the spread of several new variants. Thus, an unmet need remains for the development of novel vaccines able to target several viral strains and confer wide-spread protection in the global population. Next generation of vaccines would benefit from eliciting T-cell mediated immunity towards multiple antigens, adding a further level of protection in addition to existing humoral immunity towards the Spike glycoprotein. Here we highlight several promising antigens for which a number of broadly reactive T cell epitopes were identified following immunopeptidomics assessment of their presentation.

We assembled a comprehensive in-depth profiling of BLCL-derived immunopeptides leading to the detection of a total of 128 non-redundant SARS-CoV-2 peptides in HLA-class I elutions and 158 unique SARS-CoV-2 peptides in HLA-class II elutions of stably transfected BLCL and collectively a total of 248 unique peptides. The peptides bind highly prevalent HLA alleles expressed by the chosen BLCLs where significant gaps existed in the currently available ligand data in public databases covering HLA-B\*08:01/B\*27:05, HLA-C\*01:02/C\*07:01 to a large degree and to some extent for HLAs A\*01:01/A\*01:02. Similar additions were made for HLA-II allotypes covering HLA-DQA\*05:01/DQB1\*02:01 and HLA-DRA1\*01/HLA-DRB1\*01:01/DRB1\*03:01. Such expansion of HLA ligand coverage enables improved designs of T-cell vaccines in the future, increasing the ability to generate broad population coverages using multiple targets<sup>13</sup>.

Using peptide binding prediction tools in combination with an in vitro peptide stimulation of PBMC from convalescent donors, we validated at least eight peptides as eliciting an immune response. Our findings revealed CD4<sup>+</sup> and CD8<sup>+</sup> T cell epitopes from the Nucleoprotein that could be used in the future to elicit well-rounded immune responses facilitating not only humoral immunity via activation of



**Fig. 5 | Comparison of data generated in the current study and data available in IEDB.** Ligands identified by immunopeptidomics in the current study were compared to those deposited in the Immune Epitope Database (IEDB) for SARS-CoV-2 reading frames (A) and HLA restriction (B). The same data represented as an alignment to source proteins shows the ligand coverage of proteins analysed in the

current study and the ligand/epitope coverage from the data annotated in IEDB, divided by MHC Class I or II restriction (C). IEDB epitopes were restricted to ELI-SPOT, intracellular staining and multimer staining assay types. Nsp: non-structural protein, HLA human leukocyte antigen, MHC Major histocompatibility complex. Source data are provided as a Source Data file.

CD4<sup>+</sup> T cells but also cytotoxic and tissue-resident immunity by CD8<sup>+</sup> T cells shown to drive clearance<sup>41</sup>. Similar observations were made for E, Nsp1, and Nsp 9, adding to the pool of potential targets in future T-cell-focused vaccines. Recent vaccine development has focussed on a range of antigens such as University of Tubingen CoVac-1 (S, N, M, E, ORF8)<sup>42</sup>, EpiVax EPV-CoV19 (S, M, E), DIOSynVax (S, E, N, M), Vaxxinity UB-612 (S, M, N) and Gritstone Bio CORAL (S, N, M, ORF3a). Strikingly, without exception all peptides with confirmed T cell responses were part of detected antigen presentation hotspots (Fig. 4B). However, further studies with larger cohorts will be needed to assess whether any of the here confirmed antigens elicit immunodominant responses as this study was limited to nine donors with variable HLA allotypes. In addition to being unutilised targets to date, Nsp1 and Nsp9 are expressed early, 3 host-post-infection and 6 hpi respectively, in the viral life cycle, supporting the early clearance of infected cells, when compared to the E protein which is expressed at 24 hpi<sup>24,43</sup>. The identified peptides are highly conserved across all variants of concern (VoCs), variants of interest (VOIs) and variants being monitored (VBM) as defined by the CDC, with only a single point mutation identified for the Beta (B.1.351) variant at position E:P71L that effects peptide E:58-73.

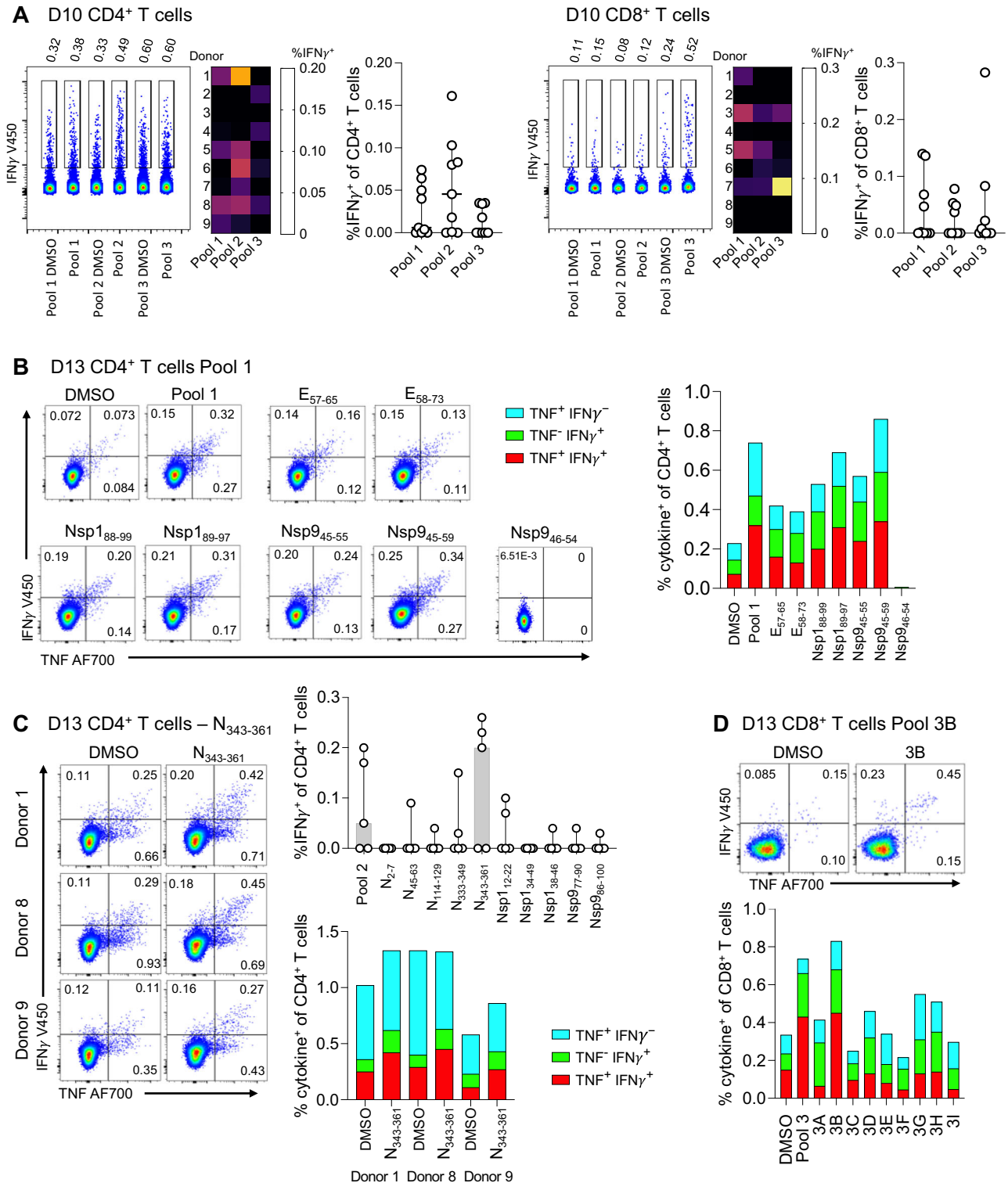
To date, it is not clear what constitutes a protein region with high antigen processing and presentation compared to other regions. More work is needed to understand how proteasome processivity, post-translational modifications, antigen structure, influence antigen

processing and why some antigenic regions go on to dominate the immunopeptidome. Overall, our study provides important insights into SARS-CoV-2-specific T-cell responses and contributes to our knowledge on experimentally verified, readily presented SARS-CoV-2 antigens. Prior to this study, the experimentally validated SARS-CoV-2 immunopeptidome was limited, but our in-depth probing of the immunopeptidome allowed us to validate 14% of the 56 tested peptides as epitopes of T-cell targets. The immunopeptidome dataset adds a further layer of information to available T cell epitope data by exposing HLA ligands that are readily processed and presented by prevalent HLA allotypes. We anticipate that this set of immunopeptides will be of importance to rationally design the next-generation COVID-19 vaccines to elicit broad T cell and B cell immunity targeting conserved epitopes across a range of emerging SARS-CoV-2 variants.

## Methods

### Cell lines and culture

The EBV transformed human B lymphoblastoid cell lines (BLCL) **IHW09004** (A\*02:01:01:01, B\*27:05:02, C\*01:02:01, DRA\*01:01, DRB1\*01:01:01, DRB6\*01:01, DQA1\*01:01:01, DQB1\*05:01:01:03, DPA1\*01:03:01:02, DPB1\*04:01:01:01) and **IHW09087** (A\*01:01:01:01, B\*08:01:01:01, C\*07:01:01:01, DRA\*01:02, DRB1\*03:01:01, DRB3\*01:01:02, DQA1\*05:01:01:02, DQB1\*02:01:01, DPA1\*01:03:01, DPB1\*03:01:01, DPB1\*04:01:01)



**Fig. 6 | CD4<sup>+</sup> and CD8<sup>+</sup> T cell reactivity towards detected SARS-CoV-2 peptides.** **A** Representative concatenated FACS plots of IFN $\gamma$  expression in Donor 1 CD4<sup>+</sup> and Donor 7 CD8<sup>+</sup> T cells in response to peptide pools 1, 2, 3 (left). Frequency of IFN $\gamma$ + CD4<sup>+</sup> and CD8<sup>+</sup> T cells across donors following stimulation with peptide pools (background DMSO subtracted). Percentage IFN $\gamma$ + of CD4<sup>+</sup> and CD8<sup>+</sup> T cells depicted as a heatmap (middle) and median with 95% CI is shown ( $n = 9$  donors, mean with SD; right). **B** Representative FACS plots of CD4<sup>+</sup> T cell IFN $\gamma$ + response (left) of Donor 3. Frequency of IFN $\gamma$ + TNF+ CD4<sup>+</sup> T cells after stimulation with individual pool 1 SARS-CoV-2 peptides (right). **C** Representative FACS plots for

CD4<sup>+</sup> T cell response to N343-361 (left). Frequency of IFN $\gamma$ + CD4<sup>+</sup> T cells stimulated with pool 2 individual peptides (background DMSO subtracted,  $n = 5$  donors, mean with SD; top). Frequency of IFN $\gamma$ + TNF+ CD4<sup>+</sup> T cells after stimulation with DMSO and N343-361 (bottom). **D** Representative FACS plot of IFN $\gamma$ + CD8<sup>+</sup> T cells (top). IFN $\gamma$ + TNF+ expression in response to stimulation with peptide pool 3 sub-pools (bottom). D10/D13 day 10/13, CD cluster of differentiation, IFN interferon, TNF Tumor necrosis factor, DMSO dimethyl sulfoxide, Nsp non-structural protein. Source data are provided as a Source Data file.



Transplantation and Immunogenetics Service. Cells were maintained in RPMI-10: RPMI (Invitrogen) supplemented with 10 % FCS, 2 mM glutamine, 1% (v/v) non-essential amino acids, 5 mM HEPES, 50  $\mu$ M  $\beta$ -mercaptoethanol, 50 IU/ml penicillin and 50  $\mu$ g/ml streptomycin in upright standing flasks at high density and a splitting regime of 1:3-1:4 in 37 °C/5 % CO<sub>2</sub>.

### Protein expression and purification

SARS-CoV-2 RNA was extracted from second passage Vero cells infected with original SARS-CoV-2 patient isolate<sup>44</sup> using the QiaAmp Viral RNA Mini Kit (Qiagen) according to manufacturer's instructions and a cDNA library was generated using the SuperScript™ IV First-Strand Synthesis System (Invitrogen) and random hexamer primers. The DNA sequence coding for SARS-CoV-2 N protein was amplified from the cDNA library via PCR using primers 5' GTAGGATCCTGTAATGGACCCCAAATCAG 3' and 5' AGTACCGGTGGCTGAGTTGAGTCAGCAC 3' and cloned into a modified pHlsec expression vector<sup>45</sup> containing a murine IgK secretion signal sequence and c-terminal TwinStrepTag. NP was expressed for 7 days in Expi293F cells (GIBCO) transiently transfected with pHlsec-NP using polyethyleneimine<sup>46</sup>. Culture supernatant was clarified at 12000 g, diluted with an equal volume of buffer containing 100 mM Tris, 150 mM NaCl 1 mM EDTA, filtered (0.8  $\mu$ M membrane), and passed over a Streptactin-XT sepharose column (IBA). Bound NP was washed extensively with BTBS buffer (20 mM Bis-Tris pH 6, 400 mM NaCl) and eluted with 20 mM Biotin BTBS. Fractions containing NP were pooled and further purified via Superdex S200 Gel permeation chromatography in BTBS. Peak fractions were concentrated using an Amicon centrifugal filter (30 kDa MWCO) and stored at -80 °C.

The coding sequences for Nsp1, 4, 5, and 9 were cloned into pET-28 vector with a cleavable N-terminal His-tag and purified as described previously for Nsp9<sup>47</sup>.

### Direct delivery of antigen

For electroporation of soluble SARS-CoV-2 derived proteins, aliquots of  $1 \times 10^7$  cells were resuspended in 0.5 ml RPMI containing 20  $\mu$ g of a purified SARS-CoV-2 protein. Following electroporation (390 V, 975  $\mu$ F,  $\infty$ ) in 4 mm MicroPulser Electroporation Cuvettes (Biorad, #1652088) using the Gene Pulser Xcell™ Electroporation Systems (Biorad), cells were transferred to tissue culture flasks and maintained in RPMI medium as above. Following incubation of 48 hrs at 37 °C/5 % CO<sub>2</sub>, cells were harvested and washed in PBS. After washing,  $0.9 \times 10^8$  to  $1.6 \times 10^8$  cells were harvested and the cell pellet snap-frozen in liquid nitrogen.

### Molecular cloning and generation of transfected cell lines

SARS-CoV-2 protein containing plasmids as deposited by the Krogan group with Addgene (141385, 141391, 141375) were used to clone genes of interest into the pEF1 $\alpha$ -IRES-DsRed-Express2 Vector (Clontech) using the EcoRI and BamHI restriction enzymes<sup>48,49</sup>. For transfection, 40  $\mu$ g of plasmid was mixed with  $1 \times 10^7$  cells in 800  $\mu$ l of RPMI medium and electroporated as above. After 48 h, cells were cultured under G418 selection and later sorted for DsRed expression. Pellets of  $4 \times 10^8$  to  $1 \times 10^9$  stably transfected cells were collected.

### Purification of peptide-HLA complexes

Cell pellets were stored at -80 °C until further use. Isolation of peptide HLA complexes has been described in detail previously<sup>16,31</sup>. Briefly, for large-scale experiments stably transfected cells were ground in a Retsch Mixer Mill MM 400 under cryogenic conditions or were directly lysed for small-scale experiments (direct antigen delivery). Cells were lysed in 0.5% IGEAL (Sigma-Aldrich, #18896), 50 mM Tris, pH 8, 150 mM NaCl (Merck-Millipore, #106404) and protease inhibitors (Complete Protease Inhibitor Cocktail Tablet, 1 tablet per 50 mL solution; Roche Molecular Biochemicals, #11697498001) for 1 hour at

4 °C with slow end-over-end mixing. Peptide-HLA complexes were immunoaffinity captured from clarified cell lysates by passing through the pan-HLA class I antibody W6/32 bound to protein A sepharose, followed by passing lysate through an HLA class II antibody mixture bound to protein A sepharose beads (anti-HLA-DQ SPV-L3: anti-HLA-DP B7/21: anti-HLA-DR LB3.1 at 1:1:1 ratio). For large-scale peptide-HLA elution, antibodies previously underwent an additional step of cross-linking to protein A sepharose. The cell lysate was co-incubated with immunoaffinity beads for at least 1 hour. Bound peptide-HLA complexes were eluted with 10% acetic acid. For small-scale peptide-HLA elution of cells undergoing direct delivery of antigen, the eluted mixture of peptides was purified by Amicon® 5 kDa Ultra-Centrifugal filter unit (Merck Millipore) and concentrated by OMIX C18 Pipette Tips (Agilent, A57003100) prior to mass spectrometric analysis. For large-scale elution, the peptide-HLA mixtures were fractionated off-line using a 4.6-mm  $\times$  100-mm monolithic reversed-phase C18 high-performance liquid chromatography (HPLC) column (Chromolith SpeedROD; Merck Millipore) and an ÄKTAmicro HPLC system (GE Healthcare). The mobile phase consisted of Buffer A (0.1% trifluoroacetic acid; Thermo Fisher Scientific) and buffer B (80% acetonitrile, 0.1% trifluoroacetic acid; Thermo Fisher Scientific). Peptide-HLA mixtures were loaded onto the column at a flow rate of 1 mL/min with separation based on a gradient of 2–40% Buffer B for 4 min, 40–45% for 4 min and a final rapid 2-min increase to 100%. Fractions (1 ml) were collected, pooled, vacuum-concentrated and diluted in 0.1% formic acid with the inclusion of retention alignment peptide standards (iRT peptides<sup>50</sup>) prior to mass spectrometric analysis.

### Mass spectrometry

Blank controls and samples listed in Figs. 2 and 3 were analysed as single injections of single replicates using a hybrid trapped ion mobility-quadrupole time of flight mass spectrometer (Bruker timsTOF Pro, Bruker Daltonics) coupled to nanoElute UHPLC liquid chromatography system. The HLA ligands were loaded onto a Trap PepMap Neo (C185mm  $\times$  300 $\mu$ m 5 $\mu$ m) trap column, eluted and separated on an IonOpticks Aurora (25 cm  $\times$  75 $\mu$ m i.d.) analytical column using a linear step-wise gradient of Buffer A (Optima water, 2% acetonitrile, 0.1% formic acid) to Buffer B (acetonitrile, 0.1% formic acid) initially 0% to 17% buffer B over 60 min, then to 25% over the next 30 min, 37% over the next 10 min followed by a rapid rise to 95% Buffer B over a subsequent 10 min period with flow rate set to 300 nL/min in PASEF mode. Data-dependent acquisition was performed with the following settings: m/z range: 100–1700 m/z, capillary voltage:1600 V, Target intensity of 30000, TIMS ramp of 0.60 to 1.60 Vs/cm<sup>2</sup> for 166 ms.

### LC-MS/MS Data Analysis

Liquid Chromatography with tandem mass spectrometry (LC-MS/MS) data was searched against the human proteome appended with the Wuhan SARS-CoV-2 proteome using PEAKS Online 10 and peptide identities subject to strict bioinformatic criteria including the use of a decoy database to apply a false discovery rate (FDR) cut-off of 5%. For SARS-CoV-2 peptides additional high confidence re-testing of detection threshold at 1% FDR was performed and 229/302 peptides (76%) were confirmed at both cut-offs (Supplementary Data S1, column H). The following search parameters were used: no cysteine alkylation, no enzyme digestion (considers all peptide bond cleavages), instrument-specific settings for TimsTOF Pro (parent and fragment ion tolerance of 20 ppm and 0.02 Da respectively), human-reviewed uniprot database (Uniprot/Swissprot v2020\_03), variable modifications set to: oxidation of Met, Acetylation of Lys and deamidation of Asn/Gln. Additionally, Peaks PTM search was performed after a Peaks DB search with all default in-built modifications with the same mass tolerance settings as Peaks DB. Cross-reactivity assessments were done using the `agrep` UNIX command ([https://github.com/PurcellLab/agrep\\_for\\_](https://github.com/PurcellLab/agrep_for_)

**crossreactivity**) to search for 1 or 2 mismatched amino acids between detected SARS-CoV-2 peptides and the human proteome (UniProt download 2022\_03). NetMHCpan and NetMHCIIpan binding prediction were used with %Rank cut-off 2 and 5 respectively to include strong and weak binders in the analysis. Synthetic peptides were ordered from Mimotopes and mirror plots were generated using Universal Spectrum Explorer (<https://www.proteomicsdb.org/use/>).

#### Data collection from IEDB, co-variants.org and bioinformatic analysis

All reported ligands and epitopes found in IEDB (<https://www.iedb.org/>, accessed 25-Feb-2024) were extracted narrowing the search to organism ID “SARS-CoV-2”, No B-cell assays and only reported data points with a defined literature reference. The resulting generated list of ligands was then manually filtered to remove any poor-quality data points e.g. only reported at the HLA I or II level or with no well-defined experimental evidence, while correcting nomenclature issues on how HLA restrictions were defined and translating the ORF1ab to separate Nsps (acc:PODTC1) using custom build python scripts. The resulting data was then plotted as a function of HLA type and gene of origin overlaying the data points reported in this manuscript. To assess the conservation of the identified reactive peptides in patient samples, all non-synonymous mutations were extracted from co-variants.org<sup>51</sup> for VOC, VOI, and VBM as defined by the CDC. A custom python script was then used to investigate any overlap between recorded mutations and the 5 genomic regions corresponding to peptides E:57-65, E:58-73, N:343-361, Nsp1:88-99 (88-99) and Nsp9:45-59 (4186-4200). Information on the SARS-CoV-2 variants coverage can be found in Supplementary Data S6.

#### Expansion of antigen-specific T cells

PBMCs were removed from storage in liquid nitrogen, thawed and washed with complete RPMI (RPMI-1640 with 10% heat-inactivated FCS, 100 mM MEM non-essential amino acids, 55 mM 2-mercaptoethanol, 5 mM HEPES buffer solution, 1 mM MEM sodium pyruvate, 1 mM L-glutamine, 100 U mL<sup>-1</sup> penicillin, and 100 mg mL<sup>-1</sup> streptomycin (Gibco/ThermoFisher Scientific)). Antigen-specific T cells were expanded essentially as previously described<sup>52</sup>, by peptide-pulsing one-third of PBMCs with a pool of up to 20 peptides at a total concentration of 10 μM for 1 h at 37 °C/5% CO<sub>2</sub>, before cells were washed twice with RPMI and added to the remaining autologous PBMCs. Cells were maintained in cRPMI at 37 °C/5% CO<sub>2</sub> for 4 days before adding and maintaining a concentration of 20 U/mL of recombinant human IL-2 (Roche Diagnostics, Mannheim, Germany).

#### T cell re-stimulation and intracellular cytokine staining

Intracellular cytokine staining was performed on days 10–13 to identify antigen-specific T cells after peptide stimulation. On day 10, T cells were restimulated with the same pool, if a response was detected, individual peptide or subpool testing was performed on day 13. T cells were restimulated with 10 μM of individual or pooled SARS-CoV-2 peptides, in the presence of brefeldin A (GolgiPlug, BD Biosciences), monensin (GolgiStop, BD Biosciences) and anti-CD107a-AF488 antibody (Invitrogen, Cat#2423749, eBioH4A3, 1:200). Cells were incubated for 5 h at 37 °C/5% CO<sub>2</sub> and then stained with anti-CD3-BV510 (Biolegend, Cat#317332, OKT3, 1:200), anti-CD4-BV650 (BD Biosciences, Cat#563875, SK3, 1:200), anti-CD8-PerCPy5.5 (BD Biosciences, Cat#565310, SK1, 1:100) and NIR Live Dead dye (Invitrogen, Cat#223869, 1:800). Cells were fixed and permeabilised using Cytofix and Cytoperm (BD Biosciences) and then stained with anti-TNF-AF700 (BD Biosciences, Cat#557996, MAb11, 1:50) and anti-IFN-γ-V450 (BD Biosciences, Cat#560371, B27, 1:100). Samples were acquired on a BD LSRII Fortessa and analysed using FlowJo v10 software.

#### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

#### Data availability

Source data are provided with this paper. The mass spectrometry data generated in this study has been deposited in the Massive repository (<https://massive.ucsd.edu/>) under data set identifier MSV000093193. Source data are provided with this paper.

#### Code availability

The agrep code is available at [https://github.com/PurcellLab/agrep\\_for\\_crossreactivity](https://github.com/PurcellLab/agrep_for_crossreactivity), <https://doi.org/10.5281/zenodo.12792072>.

#### References

- Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
- Jangra, S. et al. SARS-CoV-2 spike E484K mutation reduces antibody neutralisation. *Lancet Microbe* **2**, e283–e284 (2021).
- Low, J. S. et al. Clonal analysis of immunodominance and cross-reactivity of the CD4 T cell response to SARS-CoV-2. *Science* **372**, 1336–1341 (2021).
- Mallajosyula, V. et al. CD8(+) T cells specific for conserved coronavirus epitopes correlate with milder disease in COVID-19 patients. *Sci. Immunol.* **6**, eabg5669 (2021).
- Augusto, D. G. et al. A common allele of HLA is associated with asymptomatic SARS-CoV-2 infection. *Nature* **620**, 128–136 (2023).
- Yin, K. et al. Long COVID manifests with T cell dysregulation, inflammation and an uncoordinated adaptive immune response to SARS-CoV-2. *Nat. Immunol.* **25**, 218–225 (2024).
- Wragg, K. M. et al. Establishment and recall of SARS-CoV-2 spike epitope-specific CD4(+) T cell memory. *Nat. Immunol.* **23**, 768–780 (2022).
- Yan, L. N. et al. Neutralizing antibodies and cellular immune responses against SARS-CoV-2 sustained one and a half years after natural infection. *Front Microbiol* **12**, 803031 (2021).
- Bilich, T. et al. T cell and antibody kinetics delineate SARS-CoV-2 peptides mediating long-term immune responses in COVID-19 convalescent individuals. *Sci. Transl. Med* **13**, eabf7517 (2021).
- Nguyen, T. H. O. et al. Robust SARS-CoV-2 T cell responses with common TCRαβ motifs toward COVID-19 vaccines in patients with hematological malignancy impacting B cells. *Cell Rep. Med* **4**, 101017 (2023).
- Geers, D. et al. SARS-CoV-2 variants of concern partially escape humoral but not T-cell responses in COVID-19 convalescent donors and vaccinees. *Sci. Immunol.* **6**, eabj1750 (2021).
- Carter, B. et al. A pan-variant mRNA-LNP T cell vaccine protects HLA transgenic mice from mortality after infection with SARS-CoV-2 Beta. *Front. Immunol.* **14**, 1135815 (2023).
- Persson, G. et al. DNA immunization with in silico predicted T-cell epitopes protects against lethal SARS-CoV-2 infection in K18-hACE2 mice. *Front. Immunol.* **14**, 1166546 (2023).
- Tai, W. et al. An mRNA-based T-cell-inducing antigen strengthens COVID-19 vaccine against SARS-CoV-2 variants. *Nat. Commun.* **14**, 2962 (2023).
- Arieta, C. M. et al. The T-cell-directed vaccine BNT162b4 encoding conserved non-spike antigens protects animals from severe SARS-CoV-2 infection. *Cell* **186**, 2392–2409.e2321 (2023).
- Purcell, A. W., Ramarathnam, S. H. & Ternette, N. Mass spectrometry-based identification of MHC-bound peptides for immunopeptidomics. *Nat. Protoc.* **14**, 1687–1707 (2019).
- Brand, M. & Keşmir, C. Evolution of SARS-CoV-2-specific CD4(+) T cell epitopes. *Immunogenetics* **75**, 283–293 (2023).
- Ratishvili, T. et al. A multifaceted approach for identification, validation, and immunogenicity of naturally processed and in silico-

- predicted highly conserved SARS-CoV-2 peptides. *Vaccine* **42**, 162–174 (2024).
19. Schroeder, S. M., Nelde, A. & Walz, J. S. Viral T-cell epitopes - Identification, characterization and clinical application. *Semin. Immunol.* **66**, 101725 (2023).
  20. Tarke, A., Grifoni, A. & Sette, A. Bioinformatic and experimental analysis of T cell immune reactivity to SARS-CoV-2 and its variants. *Front Bioinform.* **2**, 876380 (2022).
  21. Prachar, M. et al. Identification and validation of 174 COVID-19 vaccine candidate epitopes reveals low performance of common epitope prediction tools. *Sci. Rep.* **10**, 20465 (2020).
  22. Parker, R. et al. Mapping the SARS-CoV-2 spike glycoprotein-derived peptidome presented by HLA class II on dendritic cells. *Cell Rep.* **35**, 109179 (2021).
  23. Knierman, M. D. et al. The human leukocyte antigen Class II Immunopeptidome of the SARS-CoV-2 Spike Glycoprotein. *Cell Rep.* **33**, 108454 (2020).
  24. Weingarten-Gabbay, S. et al. Profiling SARS-CoV-2 HLA-I peptidome reveals T cell epitopes from out-of-frame ORFs. *Cell* **184**, 3962–3980.e3917 (2021).
  25. Weingarten-Gabbay, S. et al. The HLA-II immunopeptidome of SARS-CoV-2. *Cell Rep.* **43**, 113596 (2024).
  26. Nagler, A. et al. Identification of presented SARS-CoV-2 HLA class I and HLA class II peptides using HLA peptidomics. *Cell Rep.* **35**, 109305 (2021).
  27. Pan, K. et al. Mass spectrometric identification of immunogenic SARS-CoV-2 epitopes and cognate TCRs. *Proc. Natl Acad. Sci. USA* **118**, e2111815118 (2021).
  28. Yin, S. et al. Integrated immunopeptidomic and proteomic analysis of COVID-19 lung biopsies. *Front. Immunol.* **14**, 1269335 (2023).
  29. Lancaster, A. K., Single, R. M., Solberg, O. D., Nelson, M. P. & Thomson, G. PyPop update—a software pipeline for large-scale multilocus population genomics. *Tissue Antigens* **69**, 192–197 (2007).
  30. Bojkova, D. et al. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* **583**, 469–472 (2020).
  31. Pandey, K., Ramarathinam, S. H. & Purcell, A. W. Isolation of HLA bound peptides by immunoaffinity capture and identification by mass Spectrometry. *Curr. Protoc.* **1**, e92 (2021).
  32. Reid, S. W. et al. Antagonist HIV-1 Gag peptides induce structural changes in HLA B8. *J. Exp. Med.* **184**, 2279–2286 (1996).
  33. Gfeller, D. et al. The length distribution and multiple specificity of naturally presented HLA-I ligands. *J. Immunol.* **201**, 3705–3716 (2018).
  34. Sarkizova, S. et al. A large peptidome dataset improves HLA class I epitope prediction across most of the human population. *Nat. Biotechnol.* **38**, 199–209 (2020).
  35. Ramarathinam, S. H., Ho, B. K., Dudek, N. L. & Purcell, A. W. HLA class II immunopeptidomics reveals that co-inherited HLA-alleles within an extended haplotype can improve proteome coverage for immunosurveillance. *Proteomics* **21**, e2000160 (2021).
  36. Mei, S. Immunopeptidomic analysis reveals that deamidated HLA-bound peptides arise predominantly from deglycosylated precursors. *Mol. Cell Proteomics* **19**, 1236–1247 (2020).
  37. Phulphagar, K. M. et al. Sensitive, high-throughput HLA-I and HLA-II immunopeptidomics using parallel accumulation-serial fragmentation mass spectrometry. *Mol. Cell Proteom.* **22**, 100563 (2023).
  38. Liao, H. et al. MARS an improved de novo peptide candidate selection method for non-canonical antigen target discovery in cancer. *Nat. Commun.* **15**, 661 (2024).
  39. Supekar, N. T. et al. Variable post-translational modifications of SARS-CoV-2 nucleocapsid protein. *Glycobiology* **31**, 1080–1092 (2021).
  40. Li, C. et al. Resourcing, annotating, and analysing synthetic peptides of SARS-CoV-2 for immunopeptidomics and other immunological studies. *Proteomics* **21**, e2100036 (2021).
  41. Moss, P. The T cell immune response against SARS-CoV-2. *Nat. Immunol.* **23**, 186–193 (2022).
  42. Heitmann, J. S. et al. Phase I/II trial of a peptide-based COVID-19 T-cell activator in patients with B-cell deficiency. *Nat. Commun.* **14**, 5032 (2023).
  43. Wang, L. et al. T cell immune memory after covid-19 and vaccination. *BMJ Med.* **2**, e000468 (2023).
  44. Ogando, N. S. et al. The enzymatic activity of the nsp14 Exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J. Virol.* **94**, e01246–20 (2020).
  45. Aricescu, A. R., Lu, W. & Jones, E. Y. A time- and cost-efficient system for high-level protein production in mammalian cells. *Acta Crystallogr D. Biol. Crystallogr* **62**, 1243–1250 (2006).
  46. Fang, X. T., Sehlin, D., Lannfelt, L., Syvanen, S. & Hultqvist, G. Efficient and inexpensive transient expression of multispecific multivalent antibodies in Expi293 cells. *Biol. Proced. Online* **19**, 11 (2017).
  47. Littler, D. R., Gully, B. S., Colson, R. N. & Rossjohn, J. Crystal structure of the SARS-CoV-2 non-structural Protein 9, Nsp9. *iScience* **23**, 101258 (2020).
  48. Gordon, D. E. et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **370**, eabe9403 (2020).
  49. Gordon, D. E. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **583**, 459–468 (2020).
  50. Escher, C. et al. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **12**, 1111–1121 (2012).
  51. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).
  52. Rowntree, L. C. et al. A Shared TCR Bias toward an Immunogenic EBV Epitope Dominates in HLA-B\*07:02-Expressing Individuals. *J. Immunol.* **205**, 1524–1534 (2020).
  53. Munday, P. R. et al. Immunolyser: A web-based computational pipeline for analysing and mining immunopeptidomic data. *Comput. Struct. Biotechnol. J.* **21**, 1678–1687 (2023).

## Acknowledgements

This work was supported by the Innovation Fund Denmark (0174-00039B). We thank Lauren Howson and Julie McAuley for the provision of reagents. Authors acknowledge the Monash University FlowCore and Melbourne Cytometry Platform for flow cytometry and cell sorting. Computational resources were supported by the research cloud of Australian Research Data Commons, specifically designed for high-performance research computing. A.W.P. is supported by an Australian National Health and Medical Research Council (NHMRC) Investigator Grant (2016596). C.L. was supported by an NHMRC CJ Martin ECR Fellowship (1143366). A.B. was supported by the National Psoriasis Foundation (817907) and the Rebecca L. Cooper Medical Research Foundation (PG2020775). K.K. was supported by an NHMRC L1 (#1173871), T.H.O.N. by an NHMRC EL1 (#1194036), L.C.R. by an NHMRC EL1 (#2026357) and an MRFF Award (#2016062) to K.K., T.H.O.N. and L.C.R.; J.R. supported by NHMRC Investigator grant.

## Author contributions

A.B., L.C.R., N.T., M.S.K., J.K., S.R., N.A.M., K.K., A.B.S., A.W.P. designed research. A.B., L.C.R., Z.H., K.P., Sh.R., T.H.O.N., S.C., R.A., P.T.I., N.A.M. performed research. C.L., J.P., D.R.L., G.P., M.S.K., J.R., K.E.S., N.A.M., A.B.S. contributed new reagents/analytic tools. A.B., L.C.R., Z.H., K.P., N.T., C.L., Sh.R., T.H.O.N., E.J.L., G.P., M.S.K., J.K., N.P.C., P.F., P.T.I., K.E.S., Sri R., N.A.M., K.K., A.B.S., A.W.P. analysed data. A.B. and A.W.P. wrote the initial draft, all authors were involved in the editing of the manuscript.

## Competing interests

NT, E.J.L, GP, MSK, JVK, and ABS are employed by Evaxion Biotech A/S which holds IP for identifying neoepitopes and personalized

immunotherapy. AWP is on the scientific advisory board of Evaxion Biotech A/S. AWP is on the advisory board of Bioinformatics Solutions (Canada), and Grey Wolf Therapeutics (UK) and is a co-founder of Reseptor Therapeutics (Australia). The remaining authors declare no competing interests.

## Ethics

Human blood samples were collected between 12–223 days post-SARS-CoV-2 infection in heparinised tubes. Peripheral blood mononuclear cells (PBMCs) were isolated by density-gradient centrifugation (Ficoll-Paque, Cytiva) and cryopreserved in fetal calf serum with 10% DMSO (Cat# D2650, Sigma-Aldrich). All human experimental work was conducted according to the Declaration of Helsinki principles and the Australian NHMRC Code of Practice. All blood donors provided written informed consent. Ethics approval was granted from the Human Research Ethics Committee of Melbourne Health (HREC/66341/MH-2020) and The University of Melbourne (13344 and 20782).

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-51959-6>.

**Correspondence** and requests for materials should be addressed to Anthony W. Purcell.

**Peer review information** *Nature Communications* thanks Ricardo da Silva Antunes and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024