# Landscape and Selection of Vaccine Epitopes in SARS-CoV-2 — **Source link** ↗

Christof C. Smith, Sarah Entwistle, Caryn Willis, Steven Vensko ...+16 more authors

**Institutions:** University of North Carolina at Chapel Hill, Icahn School of Medicine at Mount Sinai, University of California, Santa Cruz, University of Georgia

Related papers:

- Landscape and selection of vaccine epitopes in SARS-CoV-2.

- Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation.

- A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2

- An immunoinformatics study on the spike protein of SARS-CoV-2 revealing potential epitopes as vaccine candidates

- LBA73 The ORF1ab of SARS-CoV-2 encodes an immunodominant epitope restricted by HLA-A*01:01

## RESEARCH                                                    Open Access

# Landscape and selection of vaccine epitopes in SARS-CoV-2

Christof C. Smith[1,2†], Kelly S. Olsen[1,2†], Kaylee M. Gentry[2], Maria Sambade[2], Wolfgang Beck[1,2], Jason Garness[2], Sarah Entwistle[2], Caryn Willis[2], Steven Vensko[2], Allison Woods[1], Misha Fini[1], Brandon Carpenter[2], Eric Routh[2], Julia Kodysh[3], Timothy O'Donnell[3], Carsten Haber[4], Kirsten Heiss[4], Volker Stadler[4], Erik Garrison[5], Adam M. Sandor[2], Jenny P. Y. Ting[2,6,7,8], Jared Weiss[2,9], Krzysztof Krajewski[10], Oliver C. Grant[11], Robert J. Woods[11], Mark Heise[2,6], Benjamin G. Vincent[1,2,12,13,14*†] and Alex Rubinsteyn[2,6,12*†]

## Abstract

**Background:** Early in the pandemic, we designed a SARS-CoV-2 peptide vaccine containing epitope regions optimized for concurrent B cell, CD4[+] T cell, and CD8[+] T cell stimulation. The rationale for this design was to drive both humoral and cellular immunity with high specificity while avoiding undesired effects such as antibody-dependent enhancement (ADE).

**Methods:** We explored the set of computationally predicted SARS-CoV-2 HLA-I and HLA-II ligands, examining protein source, concurrent human/murine coverage, and population coverage. Beyond MHC affinity, T cell vaccine candidates were further refined by predicted immunogenicity, sequence conservation, source protein abundance, and coverage of high frequency HLA alleles. B cell epitope regions were chosen from linear epitope mapping studies of convalescent patient serum, followed by filtering for surface accessibility, sequence conservation, spatial localization near functional domains of the spike glycoprotein, and avoidance of glycosylation sites.

**Results:** From 58 initial candidates, three B cell epitope regions were identified. From 3730 (MHC-I) and 5045 (MHC-II) candidate ligands, 292 CD8[+] and 284 CD4[+] T cell epitopes were identified. By combining these B cell and T cell analyses, as well as a manufacturability heuristic, we proposed a set of 22 SARS-CoV-2 vaccine peptides for use in subsequent murine studies. We curated a dataset of ~ 1000 observed T cell epitopes from convalescent COVID-19 patients across eight studies, showing 8/15 recurrent epitope regions to overlap with at least one of our candidate peptides. Of the 22 candidate vaccine peptides, 16 (n = 10 T cell epitope optimized; n = 6 B cell epitope optimized) were manually selected to decrease their degree of sequence overlap and then synthesized. The immunogenicity of the synthesized vaccine peptides was validated using ELISpot and ELISA following murine vaccination. Strong T cell responses were observed in 7/10 T cell epitope optimized peptides following vaccination. Humoral responses were deficient, likely due to the unrestricted conformational space inhabited by linear vaccine peptides.

---

* Correspondence: benjamin_vincent@med.unc.edu; alex.rubinsteyn@unc.edu
[†]Christof C. Smith, Kelly S. Olsen, Benjamin G. Vincent and Alex Rubinsteyn contributed equally to this work.
[†]Christof C. Smith and Kelly S. Olsen are co-first authors.
[1]Department of Microbiology and Immunology, UNC School of Medicine, Chapel Hill, NC, USA
[2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, CB# 7295, Chapel Hill, NC 27599-7295, USA
Full list of author information is available at the end of the article

Smith *et al. Genome Medicine*        (2021) 13:101

Page 2 of 23

**Conclusions:** Overall, we find our selection process and vaccine formulation to be appropriate for identifying T cell epitopes and eliciting T cell responses against those epitopes. Further studies are needed to optimize prediction and induction of B cell responses, as well as study the protective capacity of predicted T and B cell epitopes.

**Keywords:** SARS-CoV-2, COVID-19, vaccine, T cell, B cell

## Background

SARS-CoV-2 vaccines have largely focused on generation of B cell responses to trigger production of neutralizing antibodies [1–3]. SARS-CoV-2 enters cells through interaction of the viral receptor binding domain (RBD) with angiotensin-converting enzyme 2 (ACE2) receptors, found on the surface of human nasopharyngeal, lung, and gut mucosa [4]. Neutralizing antibodies targeting the RBD and other functional domains of the SARS-CoV-2 spike protein are a major route for achieving immunity and vaccine efficacy [5–10]. When work on this study began in March 2020, little was known about the relative contribution of different adaptive immune compartments to immunity against SARS-CoV-2. Broadly, it was understood that $CD4^+$ and $CD8^+$ T cells have roles in the antiviral immune response, including against SARS-CoV-1 [11–13]. Prior studies in SARS-CoV-1 have demonstrated T cell responses against viral epitopes, with strong T cell responses correlated with generation of higher neutralizing antibody titers [13]. Unlike antibody epitopes, T cell epitopes need not be limited to accessible regions of surface proteins. In SARS-CoV-1, concurrent $CD4^+$ and $CD8^+$ activation and central memory T cell generation were induced in exposed patients, with increased Th2 cytokine polarization observed in patients with fatal disease [13]; conversely, Th1 response has been associated with less severe disease in SARS-CoV-2 [14]. Additionally, Type 1 and Type 2 immunity are not strictly synonymous with cell-mediated and humoral immunity, respectively, with Th1 polarization capable of inducing moderate antibody production [15]. Because of these considerations, most groups developing vaccines for SARS-CoV-2 have focused on promoting Th1 response due to safety concerns and demonstrated efficacy of Th1 response [16]. To this end, we deduced that vaccines targeting humoral (B cells) and cytotoxic arms ($CD8^+$ T cells) with concurrent helper signalling ($CD4^+$ T cells), delivered with adjuvants promoting Th1 polarization, may provide optimal immunity against SARS-CoV-2.

In the intervening year, many vaccine strategies for SARS-CoV-2 have demonstrated efficacy in clinical trials, including mRNA encoding of the spike glycoprotein, recombinant spike protein, adenovirus vector expressing the surface glycoprotein, as well as delivery of whole inactivated virus [2, 3, 17–25]. These strategies have proven successful at eliciting neutralizing antibody responses against conformational epitopes [26] and offer impressive protection from both infection and disease [22, 23, 27, 28]. More recently, however, concern has emerged regarding the rapid evolution [29, 30] of the virus with concomitant decrease or loss of neutralization from some novel variants [31–33]. Currently circulating variants, however, do not appear to abrogate T cell reactivity [34] and there is hope that vaccine induced T cell responses provide a second line of defense against viral infection [35, 36]. Whether future variants would also be recognized by T cell evolutionary pressure to escape T cell responses is unclear. Multi-epitope peptide vaccination is an alternative approach which targets smaller antigenic fragments of viral proteins. Peptide vaccines have historically been most successful at eliciting T cell responses [37–40] and, in certain pathogens, they have also been able to elicit neutralizing antibodies against linear epitopes [41–44]. Peptide vaccines may have a complementary role relative to existing SARS-CoV-2 vaccines due to their history of safe administration [45–48], rapid development [49, 50], and precise selection of antigenic content. A peptide vaccine can easily exclude polymorphic antigenic regions or be updated to include antigenic fragments from newly emerging variants.

We report here a design methodology for selecting SARS-CoV-2 vaccine peptides which combines linear B cell epitopes with both $CD4^+$ and $CD8^+$ T cell epitopes, as well as an evaluation of our strategy based on a murine vaccination study and a comparison with a curated dataset of published SARS-CoV-2 T cell epitopes (Fig. 1). We start with a survey of the T and B cell epitope space of SARS-CoV-2 (Fig. 2). Predicted T cell epitopes were derived from in silico predictions filtered on binding affinity and immunogenicity models generated from epitopes deposited in the Immune Epitope Database (IEDB) [51], population diversity, and source protein abundance in order to select peptides that bind common HLA alleles and are likely to generate robust $CD8^+$ and $CD4^+$ T cell activity. B cell epitope candidates were curated from linear epitope mapping studies and further filtered by accessibility, glycosylation, polymorphism, and adjacency to functional domains to identify peptides most likely to generate robust antibody responses. Given the utility of murine-adapted SARS-CoV-2 models for evaluating vaccine candidates [7, 52–54], we also identified peptides derived from viral proteins predicted to bind murine
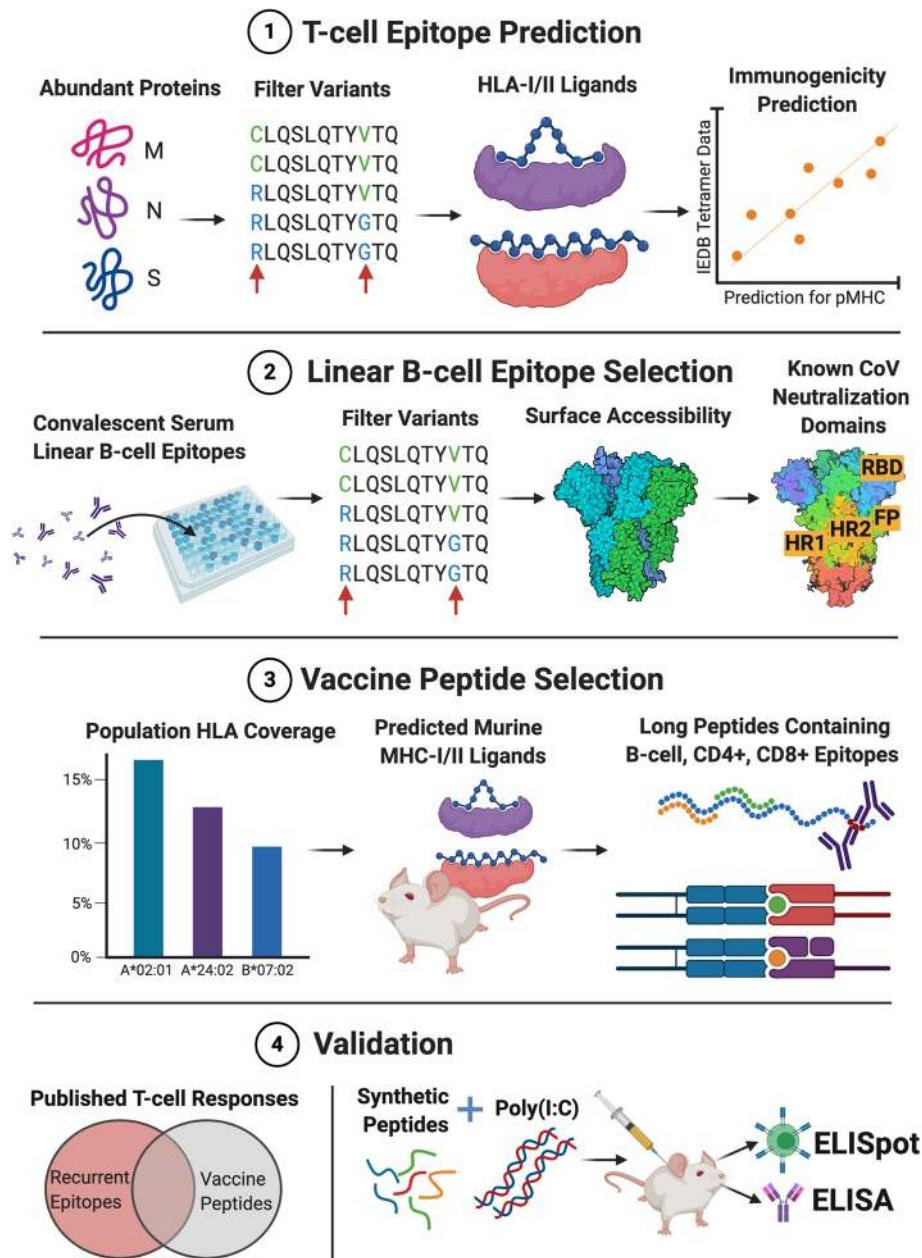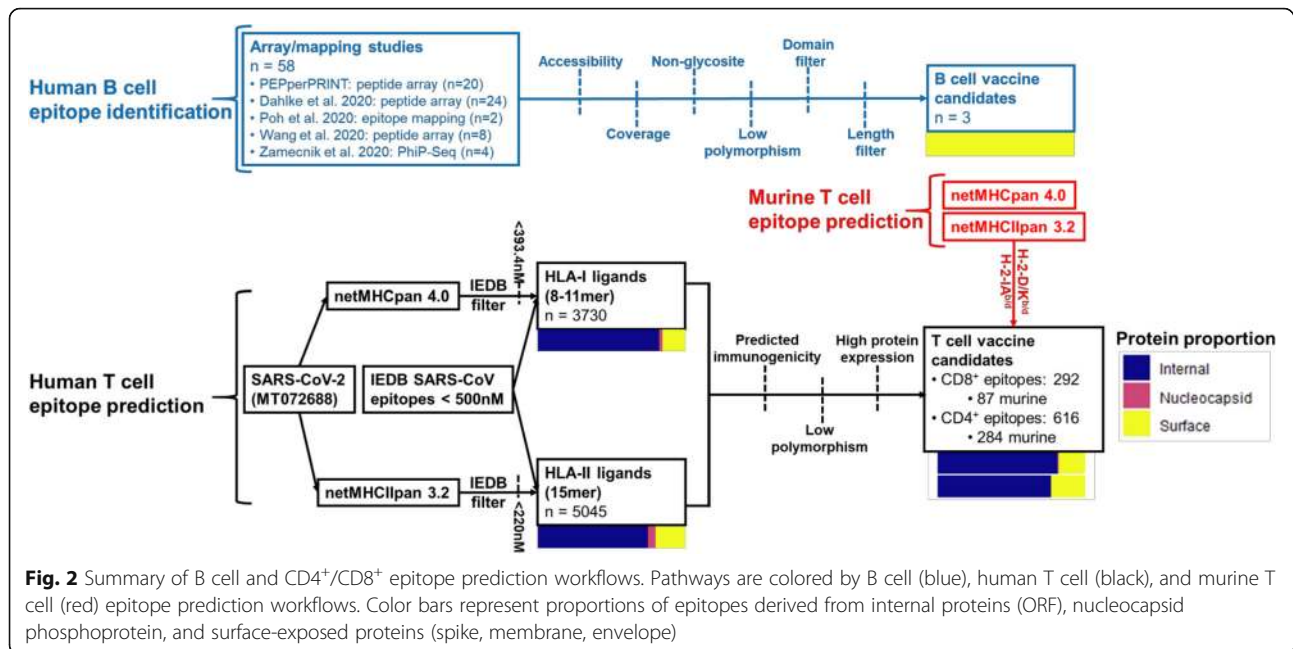
**Fig. 1** Visual summary of T and B cell epitope vaccine prediction and validation. (1) We explored the set of computationally predicted SARS-CoV-2 HLA-I and HLA-II ligands, examining source protein abundance, sequence conservation, coverage of high frequency HLA alleles, and predicted immunogenicity. (2) B cell epitope regions were chosen from linear epitope mapping studies of convalescent patient serum, followed by filtering for sequence conservation, surface accessibility, spatial localization near functional domains of the spike glycoprotein, and avoidance of glycosylation sites. (3) Vaccine selection of 27mers peptides was performed by optimizing population HLA coverage of T cell epitopes, evaluating human/murine MHC ligand co-coverage, as well as examining peptides with optimal coverage of B cell, CD4[+], and CD8[+] epitopes. (4) Lastly, validation was performed through comparison against a curated dataset of ~ 1000 observed T cell epitopes from convalescent COVID-19 patients across eight studies, as well as murine ELISA/ELISpot studies using animals vaccinated with synthetic 27mer peptides with human/murine epitope co-coverage

MHC coded for by H2-D[b/d], H2-K[b/d], and H2-IA[b/d] haplotypes. We then selected 22 longer sequence regions for use as vaccine antigens. These vaccine peptides each span multiple predicted CD4[+]/CD8[+] T cell and linear B cell epitopes, along with predicted murine MHC-I/II ligands. We compared this vaccine peptide selection process with a curated dataset of eight studies mapping SARS-CoV-2 T cell epitopes from COVID-19 patients and found that many of the recurrent epitope regions were captured by our vaccine peptides. We also

**Fig. 2** Summary of B cell and CD4+/CD8+ epitope prediction workflows. Pathways are colored by B cell (blue), human T cell (black), and murine T cell (red) epitope prediction workflows. Color bars represent proportions of epitopes derived from internal proteins (ORF), nucleocapsid phosphoprotein, and surface-exposed proteins (spike, membrane, envelope)

evaluated 16 of the 22 vaccine peptides in a murine vaccination experiment and found that the same subset of the peptides elicited T cell responses in combination with two different adjuvants.

## Methods

### Antibody epitope curation

Linear B cell epitopes on the SARS-CoV-2 surface glycoprotein were curated from five published studies [55–59]. Four of these studies screened polyclonal sera of convalescent COVID-19 patients using either peptide arrays [55, 56, 59] or phage immunoprecipitation sequencing (PhIP-Seq) [57]. One study characterized the epitopes of monoclonal neutralizing antibodies [59]. Results from Schwarz et al. included sera from six SARS-CoV-2-naive patient sera and nine SARS-CoV-2-infected patient sera using PEPperCHIP® SARS-CoV-2 Proteome Microarrays [59]. The peptides included in these proteome-wide epitope mapping analyses were limited to those which demonstrated either IgG or IgA fluorescence intensity > 1000 U in at least two infected patient samples and in none of the naive patient samples. In addition, two peptides were also included (QGQTVTKKSAAEASK, QTVTKKSAAEASKKP) which demonstrated IgG fluorescence intensity > 1000 U in only one naive patient sample each, but in four and five infected patient samples, respectively.

### HLA ligand prediction

The SARS-CoV-2 protein sequence FASTA was retrieved from the NCBI reference database (https://www.ncbi.nlm.nih.gov/nuccore/MT072688) [60]. Haplotypes

included in this analysis were derived from those with > 5% expression within the United States populations based on the National Marrow Donor Program's HaploStats tool [61]:

- HLA-A: A*11:01, A*02:01, A*01:01, A*03:01, A*24:02
- HLA-B: B*44:03, B*07:02, B*08:01, B*44:02, B*44:03, B*35:01
- HLA-C: C*03:04, C*04:01, C*05:01, C*06:02,C*07:01, C*07:02
- HLA-DR: DRB1*01:01, DRB1*03:01, DRB1*04:01, DRB1*07:01, DRB1*11:01, DRB1*13:01, DRB1*15:01
  Additionally, HLA-DQ alpha/beta pairs were chosen based on prevalence in previous studies [62]:
- HLA-DQ: DQA1*01:02/DQB1*06:02, DQA1*05:01/ DQB1*02:01, DQA1*02:01/DQB1*02:02, DQA1*05:05/DQB1*03:01, DQA1*01:01/ DQB1*05:01, DQA1*03:01/DQB1*03:02, DQA1*03:03/DQB1*03:01, DQA1*01:03/ DQB1*06:03
  For HLA-I, 8-11mer epitopes were predicted using netMHCpan 4.0 [63] and MHCflurry 1.6.0 [64]. For HLA-II calling, 15mers were predicted using NetMHCIIpan 3.2 [65] and NetMHCIIpan 4.0 [66]. For optimization of epitope predictions, individual features from each HLA-I and HLA-II prediction tool was compared against IEDB binding affinities using Spearman correlation (Additional file 1: Fig. S1). Cutpoints for the best performing HLA-I and HLA-II feature were set using 90% specificity of predicting for peptides with < 500 nM binding affinity

in the IEDB set, using predicted binding affinity values from netMHCpan 4.0 (HLA-I) and netMH-CIIpan 3.2 (HLA-II). The proportion of the total U.S. population containing at least one haplotype capable of binding each peptide was calculated assuming no genetic linkage:

$$1 - \prod_i (1 - f_i)^2$$

### Immunogenicity modeling

IEDB HLA-I and HLA-II viral tetramer data were used to generate a generalized linear model (GLM; family = binary) with tetramer-positivity as a binary outcome [51]. Independent variables for HLA-I included NetMHCpan 4.0 binding affinity and elution score, MHCflurry binding affinity, presentation score, processing score, and percentage of aromatic (F, Y, W), acidic (D, E), basic (K, R H), small (A, G, S, T, P), cyclic (P), and thiol (C, M) amino acid residues. Independent variables for HLA-II included NetMHCIIpan 4.0 binding affinity and elution scores, and percentage of aromatic, acidic, basic, small, cyclic, and thiol amino acid residues. All independent variables were normalized to 0–1 to keep coefficients comparable (binding affinities divided by 50,000). GLM model performance was derived using 5-fold cross-validation, balancing for HLA alleles. The final HLA-I and HLA-II models were generated using each full IEDB set, then applied to SARS-CoV-2 predicted HLA ligands to derive a GLM score. For immunogenicity filtering, predicted epitopes above the median GLM score were kept.

### B cell epitope selection

Accessibility of contiguous regions of the spike protein was approximated with the following heuristic: mean accessibility of 35%, minimum accessibility of 15%, requiring at least one residue to have accessibility greater than 50%, and the ends of a region to have at least 25% accessibility. Adjacency to a functional region was defined as within 15aa of either side of FP, HR1, and HR2, and within 50aa of the RBD. A broader window was used for the receptor binding domain due to the known presence of neutralizing antibody epitopes in S1 of SARS-CoV-1 outside of the RBD [67].

### Published T cell epitope data curation

T cell epitopes from eight studies of immune responses from convalescent COVID-19 patients [68–75] were manually curated into a spreadsheet with 973 entries (Table S9). Other studies were excluded which focused on murine immune responses and/or immunity from vaccination. To aggregate epitope regions of varying granularities, the viral proteome was split into 40aa bins, overlapping by 20aa. A bin was considered to contain an epitope region if they overlapped by at least 8aa. Similarly, each vaccine peptide counted as overlapping a bin if their overlap was at least 8aa. Overlapping bins were mutually exclusive, and only the bin with the highest number responding patients was retained. Bin boundaries were then clipped to the minimum and maximum boundaries of any epitope region contained within it.

### Vaccine peptide manufacturability

Based on previous experiences with peptide synthesis failures and consultation with the UNC High-Throughput Peptide Synthesis and Array Facility, we devised a scoring rubric for solid-phase peptide synthesis difficulty (Additional file 1: Fig. S8A). This rubric includes features related to the stability of the synthesized peptide product as well as sequence features which increase the difficulty of peptide elongation and/or purification. For example, hydrophobic peptides are challenging to solubilize, whereas hydrophobic regions within peptides are challenging to elongate during synthesis due to strong conformational properties. In our scoring rubric, hydrophobicity of peptide sequences is calculated using the mean GRAVY score [76], which is computed both for the entire peptide as well as the max for all local windows of lengths between 5mer and 8mer. Local hydrophobicity scores are penalized proportional to how much they exceed 2.5 whereas whole peptide hydrophobicity is penalized to the degree that it exceeds 2. These values were determined based on unpublished data relating to which peptides had failed for reasons related to hydrophobicity during the PGV001 neoantigen vaccine trial [77]. Another category of difficulties relates to the instability of certain pairs of adjacent amino acids. The extremely unstable dipeptides are DG and NG, whereas the less penalized but still problematic dipeptides are DS, DN, DD, NN, ND, NS, and NP. Furthermore, certain terminal residues inhibit the initiation of synthesis or formation of undesired residues such as pyroglutamate. Difficult N terminal residues are Q, E, C, and N, whereas difficult C terminal residues are P, C, and H. Lastly, the inclusion of multiple thiol residues can be challenging due to formation of long-range disulfide bonds. Our heuristic penalizes both the total number of thiols (C and M residues), as well as a penalty for excessive cysteines which is only applied when the number of C residues exceeds 1. Many of similar features are enumerated in commercial peptide design guides, such as ones published by Biomatik [78] and SB peptide [79] or in standard texts on solid-phase synthesis [80]. The particular weights given to different peptide features are determined purely from experience and intuition and are presented without claims of accuracy or optimality.

## SARS-CoV-2 entropy calculations

In total, 7881 SARS-CoV-2 genome sequences were downloaded from GISAID (https://www.gisaid.org/) [81]. A preprocessing step removed 127 sequences that were shorter than 25,000 bases. The sequences were split into 79 smaller files and aligned using Augur [82] (which relies on the MAFFT [83] aligner) with NCBI entry MT072688.1 [84] as the reference genome. The reference genome was downloaded from NCBI GenBank [85]. The 79 resulting alignment files were concatenated into a single alignment file with the duplicate reference genome alignments removed. The multiple sequence alignment was translated to protein space using the R packages seqinr [86] and msa [87]. Entropy for each position was calculated using the following formula, where $n$ is the number of possible outcomes (i.e., total unique identifiable amino acid residues at each location) and $p_i$ is the probability of each outcome (i.e., probability of each possible amino acid residues at each location):

$$-\sum_{i=1}^{n} p_i \cdot \log(p_i)$$

## Mouse vaccination

All mouse work was performed according to IACUC guidelines under UNC IACUC protocol ID 20-121.0. Vaccine studies were performed using BALB/c mice with free access to food and water. Mice were ordered from Jackson Laboratories and vaccinated at 8 weeks of age. Equal numbers of male and female mice were used per group, vaccinated with poly(I:C) (Sigma-Aldrich cat. #P1530) either alone or in combination with 16 synthesized vaccine peptides. In total, 26 μg total peptide was utilized per vaccination (divided equimass per peptide). Then, 75 μg of polyI:C was utilized per vaccination, with n = 6 mice per experimental group and n = 3 mice per polyI:C-only control group. Mice were vaccinated on days 1 and 7, cheek bleeds obtained on days 7 and 14, and sacrificed with cardiac bleeds performed on day 21.

## S Protein ELISA

Serum obtained from cardiac bleeds on day 21 was utilized for ELISA testing for antibody response to SARS-CoV-2 spike (S) protein. Nunc Maxisorp plates (Thermo Fisher Scientific) were coated with S protein (generously provided by Ting Lab at UNC), or BSA as a negative control and incubated overnight. Plates were blocked with 10% FBS in PBS, washed, and serum plated in duplicate wells with serial dilutions. 6x His Tagged monoclonal antibody (Thermo Fisher Scientific) was also plated as an experimental control. Goat anti-mouse IgG HRP (Thermo Fisher Scientific) was added to washed plates as a secondary antibody. TMB substrate (Thermo

Fisher Scientific) was added, development was stopped with TMB Stop solution (BioLegend), and plates were read at 450 nm.

## Peptide ELISA

Serum obtained from cardiac bleeds on day 21 and cheek bleeds on experimental days 7 and 14 were tested for antibody response to the predicted B cell peptide epitopes used for vaccinations via peptide ELISAs. Plates were coated with 5μg/mL of target peptide using coating reagent from the Takara Peptide Coating Kit (Takara cat. #MK100). Measles peptide was utilized as a negative control, and Flag peptide was also plated as an experimental control. Plates were blocked with a blocking buffer according to the manufacturer's protocol. Serum was plated in duplicate wells with serial dilutions, and anti-FLAG antibody was plated in the experimental control wells. Rabbit anti-mouse IgG HRP (Abcam ab97046) was utilized as a secondary antibody. TMB substrate (Thermo Fisher Scientific cat. #34028) was added, development was stopped with TMB Stop solution (BioLegend cat. #423001), and plates were read at 450 nm.

## ELISpot

After the sacrifice of mice on experimental day 21, spleens were dissected out for ELISpot assessment of T cell activation in response to peptide and adjuvant vaccination. Spleens were mechanically dissociated using a GentleMACS Octo Dissociator (Miltenyi Biotec) and passed through a 70-μm filter. RBC lysis buffer (Gibco cat. #A1049201) was used to remove red blood cells, and cells were washed then passed through 40-μm filters. Splenocytes were counted and 250,000 splenocytes were plated per well into plates (BD Biosciences; cat. #551083) that had been coated with each of the individual 16 predicted target peptides, or PBS as negative control or PHA as experimental control. Plates were incubated for 72 h. Anti-interferon gamma detection antibody was added according to the manufacturer's protocol, followed by enzyme conjugate Streptavidin-HRP and final substrate solution (BD Biosciences; cat. #557630). Plates were allowed to develop, washed to stop development, and allowed to dry before reading on ELISpot reader (AID Classic ERL07).

## Graphical and statistical analysis

Plots and analyses were generated using the following R packages: caret 6.0-84 [88], cowplot 0.9.4 [89], data.table 1.12.8 [90], DESeq2 1.22.2 [91], doMC 1.3.6 [92], dplyr 0.8.4 [93], forcats 0.4.0 [94], GenomicRanges 1.34.0 [95], ggallin 0.1.1 [96], ggbeeswarm 0.6.0 [97], ggnewscale 0.4.1 [98], ggplot2 3.3.0 [89], ggpubr 0.2 [99], ggrepel 0.8.1 [100], gplots 3.0.3 [101], gridExtra 2.3 [102], huxtable 4.7.1 [103], magrittr 1.5 [104], officer 0.3.10 [105],

pROC 1.16.2 [106], RColorBrewer 1.1-2 [107], readxl 1.3.1 [108], scales 1.1.0 [109], seqinr 3.6-1 [86], stringr 1.4.0 [110], venneuler 1.1-0 [111], viridis 0.5.1 [112]. Figures 4C, D and 5 were generated using the following Python packages: NumPy [113], pandas [114], Matplotlib [115], and Jupyter [116].

## Results

### Landscape of MHC ligands in SARS-CoV-2

To determine the landscape of potential HLA ligands in SARS-CoV-2 (Fig. 2, black), we first identified candidate MHC ligands by performing HLA-I binding prediction using NetMHCpan 4.0 (both EL (elution ligand) and BA (binding affinity) mode) [63] and MHCflurry [64] (8–11mers), and HLA-II binding prediction using NetMH-CIIpan 3.2 [65] and 4.0 [66] (15mers), using alleles with > 5% genetic frequency in the USA [61, 62] and worldwide populations [117] (full predicted sets for U.S. alleles: Table S1, S2; worldwide alleles: Table S3, S4). To assess the accuracy of these peptide/MHC binding prediction tools on viral peptides, we tested their performance on IEDB MHC affinity assay data values for viral peptides. Of the predictive models evaluated, NetMHC-pan 4.0 (BA) and NetMHCIIpan 3.2 demonstrated the highest correlation of binding affinity predictions for Class I and Class II MHC, respectively (Additional file 1: Fig. S1A-B). Therefore, these two predictors were used for predicting MHC ligands. A measured peptide/MHC binding affinity of 500 nM or less is commonly used to identify MHC-binding peptides which are more likely to be T cell epitopes [118, 119]. To account for the inaccuracy inherent to prediction (as opposed to measurement) of peptide-MHC affinity, we derived slightly stricter cutoffs. In order to achieve 90% specificity in IEDB binding affinity data (validated ligand set), we use predicted binding affinity thresholds of 393.4 nM and 220.0 nM for Class I and Class II MHC, respectively (Additional file 1: Fig. 1C-D). This filter was applied to NetMHCpan 4.0 and NetMHCIIpan 3.2 SARS-CoV-2 MHC binding predictions, which removed the majority of viral protein subsequences (Additional file 1: Fig. 2A-B).

After filtering by binding affinity, we observed a total of 2486 unique HLA-I ligands and 3138 unique HLA-II ligands (Fig. 3C). Predicted MHC ligands were not evenly distributed across the proteome, with local peaks and troughs observed that correlated between HLA-I and HLA-II ligands (Fig. 3C, bottom; Pearson correlation of HLA-I/II LOESS, r = 0.703, p < 0.001). Notably, while SARS-CoV-1 T cell epitopes previously described in the literature were primarily located in the surface glycoprotein (S) and nucleocapsid protein (N) (Table S5) [13, 120–145], we observed a paucity of predicted MHC ligands in the N protein. As murine models for SARS-CoV-2 would be a powerful tool in understanding viral

immunobiology, we determined which predicted HLA ligands were also predicted to bind murine MHC alleles of the H2[b] and H2[d] haplotypes. NetMHCpan and NetMHCIIpan were run using the SARS-CoV-2 proteome against the H2[b] and H2[d] haplotypes, filtering by MHC-I ligands in the top 2nd percentile (n = 3053) and MHC-II ligands in the top 10th percentile (n = 1648). From this set, we observed an overlap of 887 peptides in MHC-I and 1571 peptides in MHC-II between murine and human sets (Fig. 3D). For the nested HLA ligand set, we observed 825 and 848 overlapping murine MHC-I and MHC-II ligands, respectively, with 846 HLA ligands containing both murine MHC-I and MHC-II coverage. The majority of HLA ligand sequences were predicted to bind to fewer than 50% of the U.S. population, particularly for HLA-I ligands (Fig. 3E). In accordance with higher population coverage distribution in HLA-II, predicted HLA-II ligands also demonstrated more binding alleles on average (mean alleles per peptide: HLA-I = 1.35, HLA-II = 2.80). Among the most common alleles were HLA-A*02:01 (n = 784), HLA-A*11:01 (n = 643), and HLA-A*03:01 (n = 383) for predicted HLA-I binding peptides and HLA-DRB1*01:01 (n = 5401), HLA-DRB1*07:01 (n = 3225), and HLA-DRB1*13:01 (n = 3022) for predicted HLA-II binding peptides.

### CD8[+] and CD4[+] T cell epitope prediction

Peptide/MHC binding is necessary but not sufficient for peptide epitopes to elicit T cell responses. We sought to identify a set of epitopes that would serve as good targets for a SARS-CoV-2 T cell vaccine. From the total pool of HLA-I, HLA-II, and nested MHC ligands, we sought to prioritize sequences which are predicted to be immunogenic from highly conserved regions of abundant viral proteins (Fig. 4, middle).

To predict the immunogenicity of MHC ligands, we fit a forward stepwise multivariable logistic regression model using peptide/HLA tetramer flow cytometry data curated from viral entries of the IEDB [51]. Tetramer data was selected for the response variable because it provides unambiguous association between a peptide and its bound MHC, and additionally tests which specific peptide/MHC is capable of eliciting a T cell response. Each unique peptide-MHC was encoded with features derived from epitope prediction tools as well as features relating to amino acid content (see "Immunogenicity modeling"). Epitope prediction tool features were selected to allow for consideration of predicted binding affinity alongside other tangential features such as MHC ligand elution (NetMHCpan 4.0, NetMHCIIpan 4.0) and antigen processing (MHCflurry), while amino acid content was considered due to prior studies demonstrating capacity of these features to predict for epitope
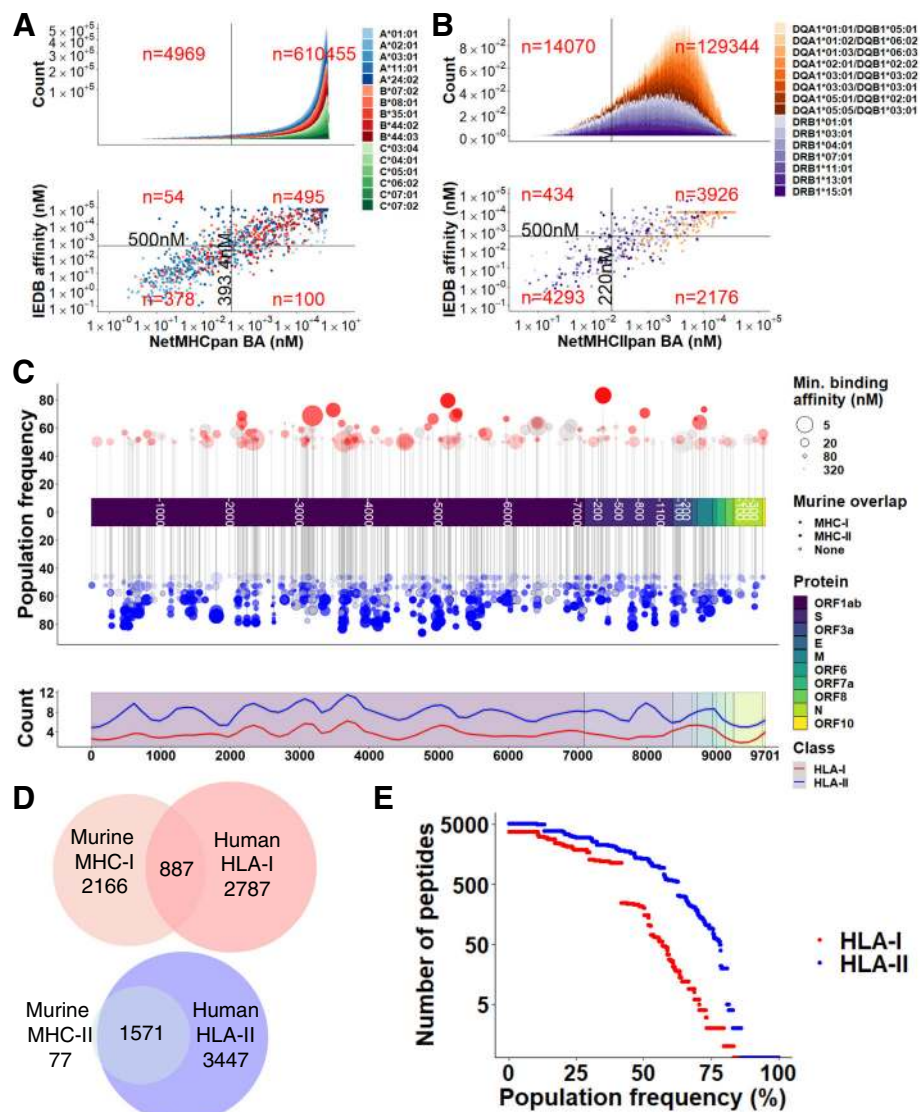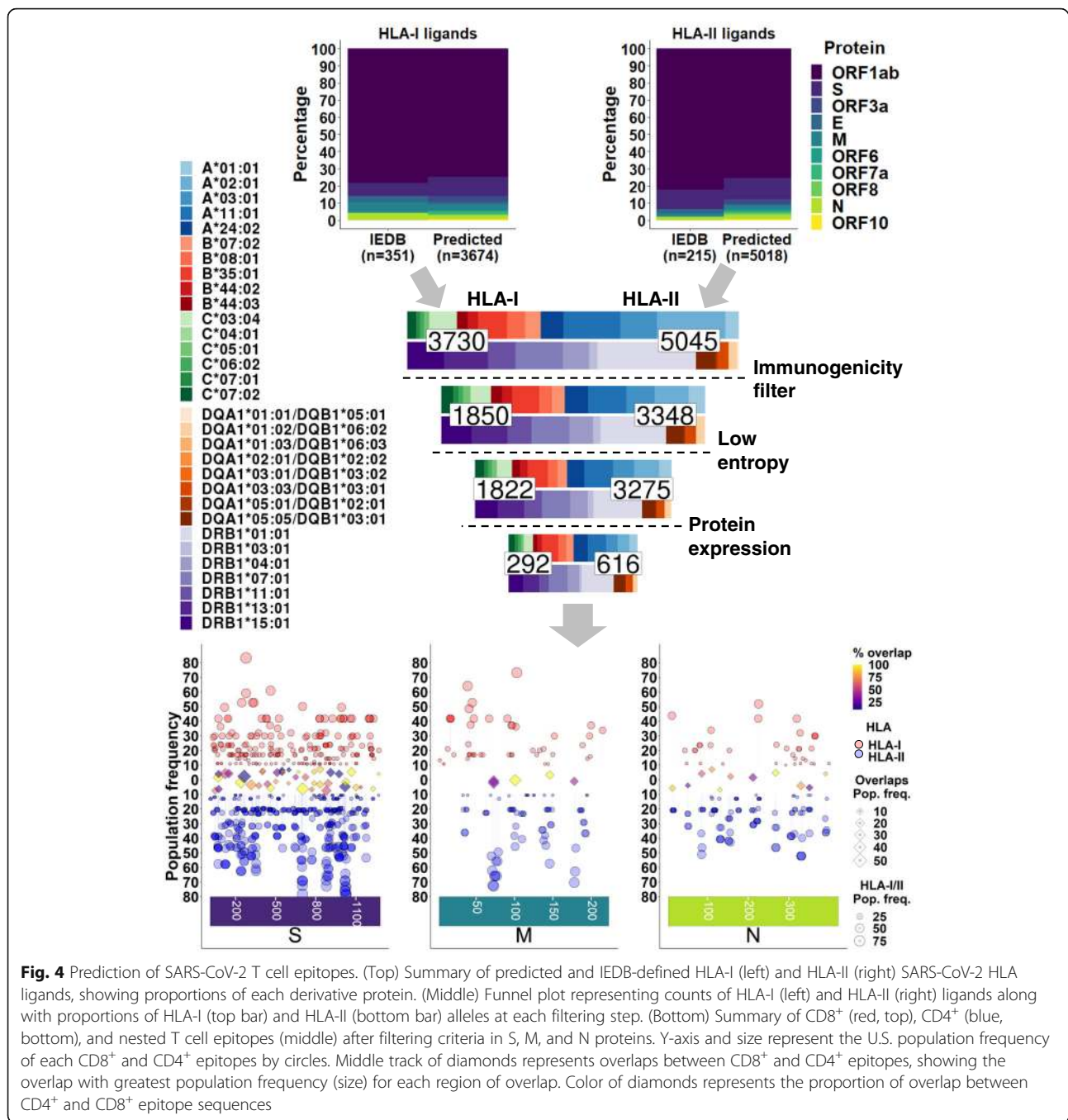
**Fig. 3** Landscape of SARS-CoV-2 MHC ligands. **A,B** Selection criteria for **A** HLA-I and **B** HLA-II SARS-CoV-2 HLA ligand candidates. Scatterplot (bottom) shows predicted (x-axis) versus IEDB (y-axis) binding affinity, with horizontal line representing 500 nM IEDB binding affinity and vertical line representing corresponding predicted binding affinity for 90% specificity in binding prediction. Histogram (top) shows all predicted SARS-CoV-2 HLA ligand candidates. Scatterplot in **B** shows subsampled points from HLA-DRB1 alleles (< 50 points per allele) to allow for increased visibility of points. **C** Landscape of predicted HLA ligands, showing HLA-I (red) and HLA-II (blue) ligands with U.S. population coverage > 50% (top), and LOESS fitted curve (span = 0.1) for HLA-I/II ligands by location along the SARS-CoV2 proteome (color tracks). The predicted binding affinity of HLA ligand peptides to murine H2-b/d alleles is represented with point shading. **D** Summary of murine/human MHC ligand overlap. **E** Distribution of population frequencies among predicted HLA-I and HLA-II ligands

immunogenicity [146, 147]. Model performance in 5-fold cross-validation demonstrated AUC values of approximately 0.7 and 0.9 for HLA-I and HLA-II, respectively, in both training and test sets Additional file 1: Fig. S2A-B). Models demonstrated cleaner separation of tetramer positive and negative groups for CD4[+] epitopes compared to CD8[+] (Additional file 1: Fig. S2C-D). To determine a cause for this difference in model performance, we examined predicted binding affinity scores

between tetramer positive and negative epitopes, which demonstrated significantly better separation for CD4[+] epitopes than CD8[+] epitopes (Additional file 1: Fig. S2E-F). In accordance with this difference in binding affinity distribution, the HLA-II model showed strong association between lower binding affinity and lower predicted tetramer positivity, while the HLA-I model showed a weaker inverse association (Additional file 1: Fig. S3). Due to these binding affinity distribution differences

**Fig. 4** Prediction of SARS-CoV-2 T cell epitopes. (Top) Summary of predicted and IEDB-defined HLA-I (left) and HLA-II (right) SARS-CoV-2 HLA ligands, showing proportions of each derivative protein. (Middle) Funnel plot representing counts of HLA-I (left) and HLA-II (right) ligands along with proportions of HLA-I (top bar) and HLA-II (bottom bar) alleles at each filtering step. (Bottom) Summary of CD8[+] (red, top), CD4[+] (blue, bottom), and nested T cell epitopes (middle) after filtering criteria in S, M, and N proteins. Y-axis and size represent the U.S. population frequency of each CD8[+] and CD4[+] epitopes by circles. Middle track of diamonds represents overlaps between CD8[+] and CD4[+] epitopes, showing the overlap with greatest population frequency (size) for each region of overlap. Color of diamonds represents the proportion of overlap between CD4[+] and CD8[+] epitope sequences

between IEDB HLA-I and HLA-II tetramer sets, a performance-based cutoff did not allow for equal filtering of CD4[+] and CD8[+] epitopes. Therefore, we filtered by generalized linear model (GLM) predicted immunogenicity scores above the median in each HLA-I/II SARS-CoV-2 epitope group, which provided balanced selection while removing predicted low-immunogenicity epitopes (Additional file 1: Fig. S4).

Next, we sought to prioritize epitopes derived from regions of low sequence variation across viral strains. A position-based entropy filter was applied to all epitopes (Additional file 1: Fig. S5), keeping those with an entropy score ≤ 0.1 (~ 98% sequence identity, n = 7881) in all amino acid positions across MSA-aligned SARS-CoV-2 genomes downloaded from the GISAID database [81, 82]. High entropy was observed in the well-described spike protein D614G polymorphic site (Additional file 1: Fig. S5A, red dot). Other areas of high entropy included positions 3606, 4715, 5828, and 5865 of ORF1ab, and position 84 of ORF8 (all with entropy > 0.4). The

Smith *et al. Genome Medicine*      (2021) 13:101

Page 10 of 23

majority of positions demonstrated > 95% sequence identity, suggesting high homology between different SARS-CoV-2 viral genomes (Additional file 1: Fig. S5B). Lastly, as the likelihood of MHC presentation is correlated with protein expression [148], we filtered epitopes to those derived from the S, M, and N proteins. These were the three highest expressed proteins based on a semi-quantitative mass spectrometry analysis of SARS-CoV-2 protein expression (PSM count/protein length; Additional file 1: Fig. S6A) [149]. This protein abundance estimation closely matched expression levels derived from SARS-CoV-2 RNA-seq data (Additional file 1: Fig. S6B) [150]. After all these filtering steps, 292 CD8+, 616 CD4+, and 423 nested T cell epitopes were predicted. We cross-filtered these epitopes against a reference peptidome of 8-11mer and 15mer peptides derived from the GRCh38 reference proteome [151] and observed no overlap. Relative proportions of HLA alleles were conserved throughout filtering (Fig. 4, middle). Full peptide sets with all filtering criteria are listed in Tables S1 (HLA-I) and S2 (HLA-II).

## B cell epitope prediction

In addition to identifying SARS-CoV-2 T cell epitopes, we sought to identify a set of linear B cell epitopes on the spike protein which would serve as good targets for stimulating neutralizing antibody responses (Fig. 2). Epitope candidates were derived from four published preprint mapping/array studies [55, 56, 58, 59] including a PEPperCHIP® peptide array study [59] (for study details see "Antibody epitope curation"). Starting with an initial candidate pool of 58 linear epitopes with data to support in vivo generation in humans (Fig. 5A, Table S6), we applied a set of filtering criteria to narrow our target space (Fig. 5B):

1. Contiguous sub-sequences of the spike protein with high accessibility
2. Exclude glycosylation sites
3. Exclude regions with significant polymorphism between SARS-CoV-2 strains
4. Keep candidate epitopes within or adjacent to functional domains with evidence of antibody-mediated viral neutralization in SARS-CoV-1 (receptor binding domain, fusion peptide, heptad repeat regions)
5. Exclude any candidates shorter than four amino acids

We used SARS-CoV-2 S protein accessibility data from Grant et. al. [152], which calculates accessibility from molecular dynamics simulations of a spike protein structure with several different glycosylation patterns. Unfortunately, this accessibility data lacks HR2, causing

that domain to be left out from subsequent analyses. After filtering for contiguously accessible regions, there were 19 remaining under consideration. Since many epitopes occur in multiple sources, we combined overlapping epitope candidates into 14 unique sequences. After filtering out epitopes containing glycosites, which may alter antibody binding characteristics [153, 154], 11 non-glycosylated regions remained. Two additional regions were removed because they contained polymorphic sites, defined by mutation frequency > 0.1% from GISAID SARS-CoV-2 viral sequences. Of the remaining 9 regions, only 4 were close to functional domains which in the closely related virus SARS-CoV-1 have evidence of antibody-mediated viral neutralization: the RBD, fusion protein (FP), and heptad repeats [155–160]. This filtering resulted in four remaining regions, of which our final criteria removed one which had length less than four residues (Fig. 5B). This filtering criteria precluded the vast majority of total spike protein regions (Fig. 5C), with three predicted antibody binding regions (residue lengths 18, 4, and 4) remaining (Fig. 5D). All three epitope candidate regions were present on solvent-exposed surfaces of the S protein trimer 3D structure (Fig. 5E). It is worth noting that the largest region, residues 456-473 within the receptor binding motif (RBM) loop, is only accessible when the RBD is in the "open" conformation.

## Selection of human and murine SARS-CoV-2 vaccine peptides

With the above filters applied to predicted T and B cell epitope candidates, we derived a minimal collection of long vaccine peptides for all combinations of the following immunological criteria: CD4+ responses, CD8+ responses, coverage of predicted B cell epitopes, along with optional inclusion of predicted murine MHC ligands. A 27mer sequence for each vaccine peptide was selected to maximize U.S. population coverage of T cell epitopes within a peptide set, with or without additional coverage for murine H2b, H2d, or both haplotypes (Fig. 6A-B; Additional file 1: Fig. S7). If population coverage was identical for multiple candidates, peptides were also optimized based on a manufacturability difficulty scoring system (Additional file 1: Fig. S8). The peptide sequence length was inspired by previous work in cancer neoantigen vaccination [161–163] which has demonstrated strong CD8+ and CD4+ responses using 27mer peptides. Optimizing for CD4+ epitope population coverage demonstrated 88.5% population frequency encompassed by three 27mer peptides (Fig. 6B: 1, 9, and 15), while CD8+ epitope optimization provided 95.8% population frequency coverage by three 27mer peptides (Fig. 6B: 1, 4, and 14). CD4+/CD8+ co-optimization provided the best overall population coverage at 81.6% population frequency with four 27mer peptides (Fig. 6B:

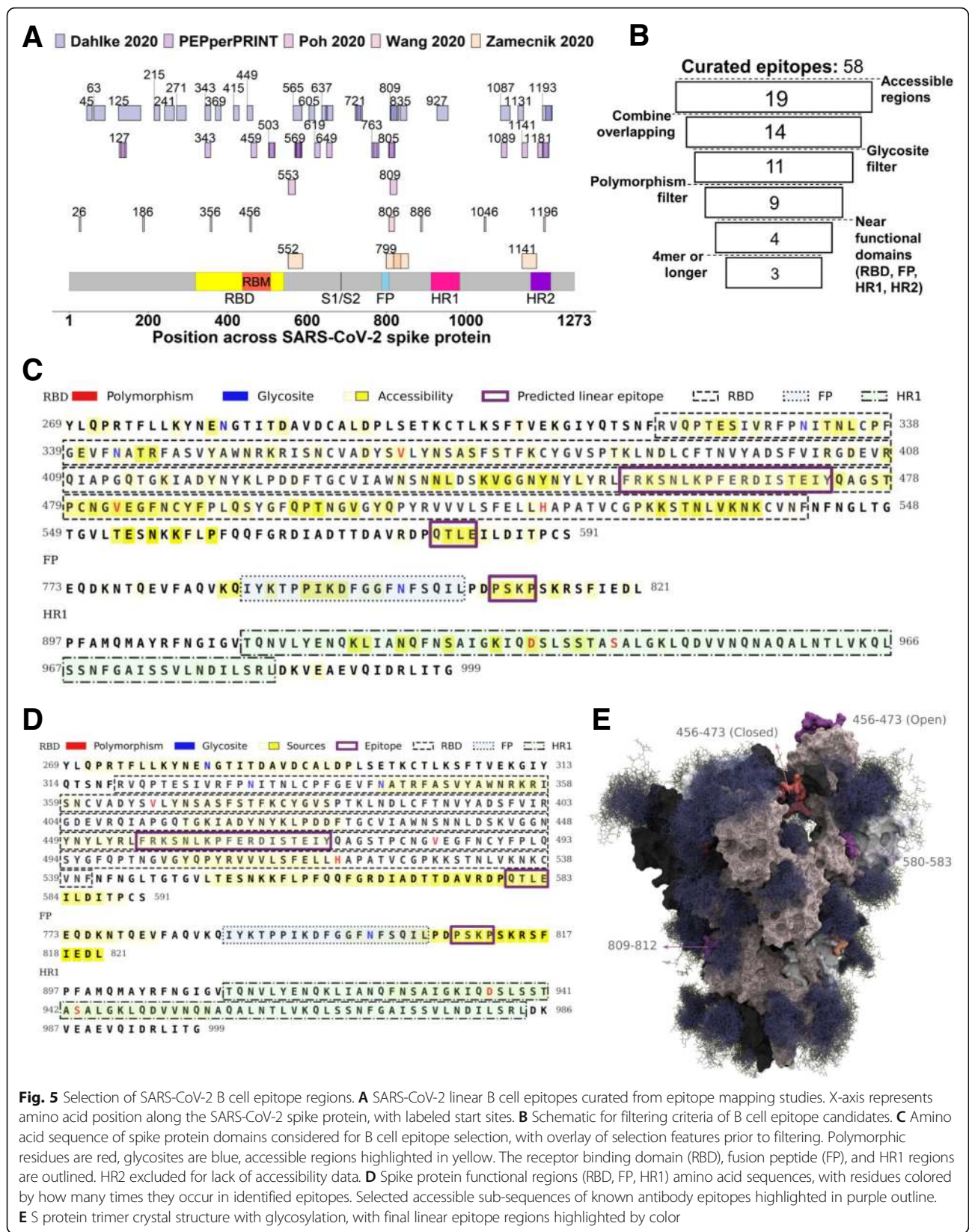Smith *et al. Genome Medicine*      (2021) 13:101

Page 11 of 23



**Fig. 5** Selection of SARS-CoV-2 B cell epitope regions. **A** SARS-CoV-2 linear B cell epitopes curated from epitope mapping studies. X-axis represents amino acid position along the SARS-CoV-2 spike protein, with labeled start sites. **B** Schematic for filtering criteria of B cell epitope candidates. **C** Amino acid sequence of spike protein domains considered for B cell epitope selection, with overlay of selection features prior to filtering. Polymorphic residues are red, glycosites are blue, accessible regions highlighted in yellow. The receptor binding domain (RBD), fusion peptide (FP), and HR1 regions are outlined. HR2 excluded for lack of accessibility data. **D** Spike protein functional regions (RBD, FP, HR1) amino acid sequences, with residues colored by how many times they occur in identified epitopes. Selected accessible sub-sequences of known antibody epitopes highlighted in purple outline. **E** S protein trimer crystal structure with glycosylation, with final linear epitope regions highlighted by color

## A

| Symbol | Set | # Peptides | HLA-I Coverage | HLA-II Coverage | Total Coverage | # B-cell Epitope Regions |
|---|---|---|---|---|---|---|
| ⊛ | CD4+/CD8+ | 4 | 92.2% | 88.5% | 81.6% | 0 |
| ⊛$^d$ | CD4+/CD8+ (H2$^d$ ligands) | 4 | 93.8% | 84.7% | 79.5% | 0 |
| ⊛$^b$ | CD4+/CD8+ (H2$^b$ ligands) | 3 | 92.2% | 84.7% | 78.1% | 0 |
| ⊛$^{bd}$ | CD4+/CD8+ (H2$^b$ and H2$^d$ ligands) | 4 | 92.1% | 84.7% | 78.0% | 0 |
| ○ | CD4+ | 3 | 91.3% | 88.5% | 80.8% | 0 |
| ○$^d$ | CD4+ (H2$^d$ ligands) | 3 | 91.3% | 88.5% | 80.8% | 0 |
| ○$^b$ | CD4+ (H2$^b$ ligands) | 3 | 76.8% | 84.7% | 65.0% | 0 |
| ○$^{bd}$ | CD4+ (H2$^b$ and H2$^d$ ligands) | 3 | 92.2% | 84.7% | 78.1% | 0 |
| * | CD8+ | 3 | 95.8% | 61.3% | 58.7% | 0 |
| *$^d$ | CD8+ (H2$^d$ ligands) | 3 | 95.1% | 76.2% | 72.5% | 0 |
| *$^b$ | CD8+ (H2$^b$ ligands) | 3 | 95.8% | 61.3% | 58.7% | 0 |
| *$^{bd}$ | CD8+ (H2$^b$ and H2$^d$ ligands) | 3 | 94.7% | 72.6% | 68.8% | 0 |
| ⊠ | B-Cell/CD4+/CD8+ | 3 | 88.9% | 62.7% | 55.7% | 3 |
| ⊡ | B-Cell/CD4+ | 3 | 88.9% | 62.7% | 55.7% | 3 |
| ⊡$^d$ | B-Cell/CD4+ (H2$^d$ ligands) | 1 | 66.2% | 39.9% | 26.4% | 1 |
| ⊡$^b$ | B-Cell/CD4+ (H2$^b$ ligands) | 2 | 64.8% | 39.4% | 25.5% | 2 |
| ⊞ | B-Cell/CD8+ | 3 | 90.8% | 57.7% | 52.4% | 3 |
| ⊞$^d$ | B-Cell/CD8+ (H2$^d$ ligands) | 1 | 81.8% | 38.4% | 31.4% | 1 |
| ⊞$^b$ | B-Cell/CD8+ (H2$^b$ ligands) | 2 | 89.4% | 46.5% | 41.5% | 2 |
| ⊞$^{bd}$ | B-Cell/CD8+ (H2$^b$ and H2$^d$ ligands) | 1 | 81.8% | 38.4% | 31.4% | 1 |
| □ | B-Cell | 3 | 81.8% | 52.8% | 43.2% | 3 |

## B

| Sequence | Protein | Start | End | B-cell Epitope | HLA-I Coverage | HLA-II Coverage | H2$^b$ I | H2$^b$ II | H2$^d$ I | H2$^d$ II | Selection Sets |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LLQFAYANRNRFLYIIKLIFLWLLWPV | M | 34 | 60 | | 89.0% | 36.0% | + | + | + | + | * *$^b$ *$^d$ *$^{bd}$ ○ ○$^d$ ○$^{bd}$ ⊛ ⊛$^d$ ⊛$^b$ ⊛$^{bd}$ |
| PVTLACFVLAAVYRINWITGGIAIAMA | M | 59 | 85 | | 42.0% | 76.0% | + | + | − | + | ○$^b$ |
| YFIASFRLFARTRSMWSFNPETNILLN | M | 95 | 121 | | 78.0% | 53.0% | + | + | + | + | ⊛$^{bd}$ |
| KDLSPRWYFYYLGTGPEAGLPYGANKD | N | 102 | 128 | | 49.0% | 39.0% | + | + | + | − | * *$^b$ *$^d$ |
| WPQIAQFAPSASAFFGMSRIGMEVTPS | N | 301 | 327 | | 63.0% | 61.0% | + | + | + | + | ○$^{bd}$ ⊛$^d$ ⊛$^{bd}$ |
| AQFAPSASAFFGMSRIGMEVTPSGTWL | N | 305 | 331 | | 71.0% | 57.0% | + | + | + | − | ⊛ ⊛$^b$ |
| SASAFFGMSRIGMEVTPSGTWLTYTGA | N | 310 | 336 | | 76.0% | 45.0% | + | − | + | − | *$^{bd}$ |
| VTPSGTWLTYTGAIKLDDKDPNFKDQV | N | 324 | 350 | | 50.0% | 62.0% | + | + | − | − | ○$^b$ |
| PQRQKKQQTVTLLPAADLDDFSKQLQQ | N | 383 | 409 | | 11.0% | 52.0% | − | − | − | + | ○ ○$^d$ ⊛ |
| YPDKVFRSSVLHSTQDLFLPFFSNVTW | S | 38 | 64 | | 44.0% | 52.0% | − | + | + | + | ⊛$^d$ |
| GAAAYYVGYLQPRTFLLKYNENGTITD | S | 261 | 287 | | 88.0% | 38.0% | + | + | + | − | *$^{bd}$ |
| SETKCTLKSFTVEKGIYQTSNFRVQPT | S | 297 | 323 | | 54.0% | 52.0% | − | − | + | − | *$^d$ |
| GLTVLPPLLTDEMIAQYTSALLAGTIT | S | 857 | 883 | | 66.0% | 73.0% | + | + | + | + | ⊛ ⊛$^d$ ⊛$^b$ ⊛$^{bd}$ |
| SVLNDILSRLDKVEAEVQIDRLITGRL | S | 975 | 1001 | | 72.0% | 28.0% | + | − | − | − | * *$^b$ |
| RLQSLQTYVTQQLIRAAEIRASANLAA | S | 1000 | 1026 | | 54.0% | 81.0% | − | + | + | + | ○ ○$^d$ ○$^b$ ○$^{bd}$ |
| GNYNYLYRLFRKSNLKPFERDISTEIY | S | 447 | 473 | S$_{456-473}$ | 82.0% | 38.0% | + | − | + | − | ⊞ ⊞$^d$ ⊞$^b$ ⊞$^{bd}$ |
| YLYRLFRKSNLKPFERDISTEIYQAGS | S | 451 | 477 | S$_{456-473}$ | 78.0% | 46.0% | + | − | − | − | □ ⊡ ⊠ |
| FRKSNLKPFERDISTEIYQAGSTPCNG | S | 456 | 482 | S$_{456-473}$ | 46.0% | 30.0% | − | + | − | − | ⊡$^b$ |
| KFLPFQQFGRDIADTTDAVRDPQTLEI | S | 558 | 584 | S$_{580-583}$ | 0.0% | 0.0% | − | − | − | − | □ |
| PQTLEILDITPCSFGGVSVITPGTNTS | S | 579 | 605 | S$_{580-583}$ | 13.0% | 21.0% | − | − | − | − | ⊞ ⊡ ⊠ |
| IYKTPPIKDFGGFNFSQILPDPSKPSK | S | 788 | 814 | S$_{809-812}$ | 35.0% | 23.0% | − | + | − | − | □ ⊡$^b$ |
| PSKPSKRSFIEDLLFNKVTLADAGFIK | S | 809 | 835 | S$_{809-812}$ | 66.0% | 40.0% | + | − | − | + | ⊞ ⊞$^b$ ○ ○$^d$ ⊠ |

**Fig. 6** T cell and B cell vaccine candidates. **A** 27mer vaccine peptide sets selecting for best CD4[+], CD8[+], CD4[+]/CD8[+], and B cell epitopes with HLA-I, HLA-II, and total U.S. population coverage. **B** Unified list of all selected 27mer vaccine peptides. Vaccine peptides containing predicted ligands for murine MHC alleles (H2-b and H2-d haplotypes) are indicated in their respective columns

1, 6, 9, 13). While B cell epitope optimization provided CD8[+] coverage above 85%, CD4[+] coverage was only 52.8%, suggesting the design of a combination B cell/ CD4[+] T cell vaccine requires use of non-spatially overlapping sequences. Overall, selection of peptides which also provided both H2[b] and H2[d] epitope coverage did not greatly impact population coverage, suggesting these murine-encompassing sets may allow for vaccine studies in animal models whilst preserving human relevance. Across the different selection criteria for minimal vaccine peptide sets, there was significant redundancy. Collapsing the set of vaccine peptides by unique sequences results in a final set of 22 27mer vaccine peptides (Fig. 6B). In addition to 27mer peptides, all individual T/ B cell epitopes (S, M, and N: Table S7; all proteins: Table S8) as well as 15mer (Additional file 1: Fig. 9) and

21mer (Additional file 1: Fig. S10) optimized peptide sets are also available.

## Validation of T cell predictions by comparison with recurrent published T cell epitopes from COVID-19 patients

To determine how our predictions of CD8+ and CD4+ T cell epitopes relate to actual SARS-CoV-2 T cell epitopes, we curated a dataset of published T cell epitope mapping studies (Table S9) and compared recurrent epitope regions with vaccine peptides. We focused on human studies of infection induced immunity, excluding murine and vaccine studies, as well as excluding studies which only performed TCR sequencing. We were able to curate eight diverse studies [68–74, 164] whose study characteristics are summarized in Fig. 7A. The T cell response assays included ELISpot, MHC multimers, MIRA [165], AIM [166], and T-Scan [167]. It is important to note that not all studies examined responses to the same proteins or even the same peptides within a protein. Some studies conducted exhaustive unbiased tiling over the viral proteome [68, 72, 74, 164], while others used computational predictions of MHC affinity to select small sets of peptides [68–71, 73]. Of these studies, only those which used multimeric MHC assays were able to unambiguously identify biological HLA restriction and the exact peptide determinants of a T cell response, whereas others used predicted or statistical assignments, sometimes within large peptide windows. To overcome the heterogeneity of this dataset, we binned the viral proteome into regions of 40 amino acids into which each study could contribute one or more identified epitope regions. A small number of recurrent epitope regions contained responses from three or more studies (Fig. 7B). Inspection of these recurrent regions broadly confirms the choice of S and N as particularly immunogenic proteins, likely due to their abundance, as well as one recurrent epitope region in the M protein. We also see strong recurrent responses to two regions of ORF3a, as well as three regions within non-structural proteins contained within ORF1ab (nsp3, nsp12, nsp13), which were not selected for consideration in our study. The identified recurrent epitope regions were strongly enriched for overlap with vaccine peptides selected in this study. In fact, 8/15 recurrent epitope regions in the S, M, and N proteins (and 8/20 total recurrent epitope regions) significantly overlapped at least one vaccine peptide. This degree of concordance gives us confidence that our computational selection process for T cell epitopes is at least to some degree predictive of biological SARS-CoV-2 T cell epitopes following infection.

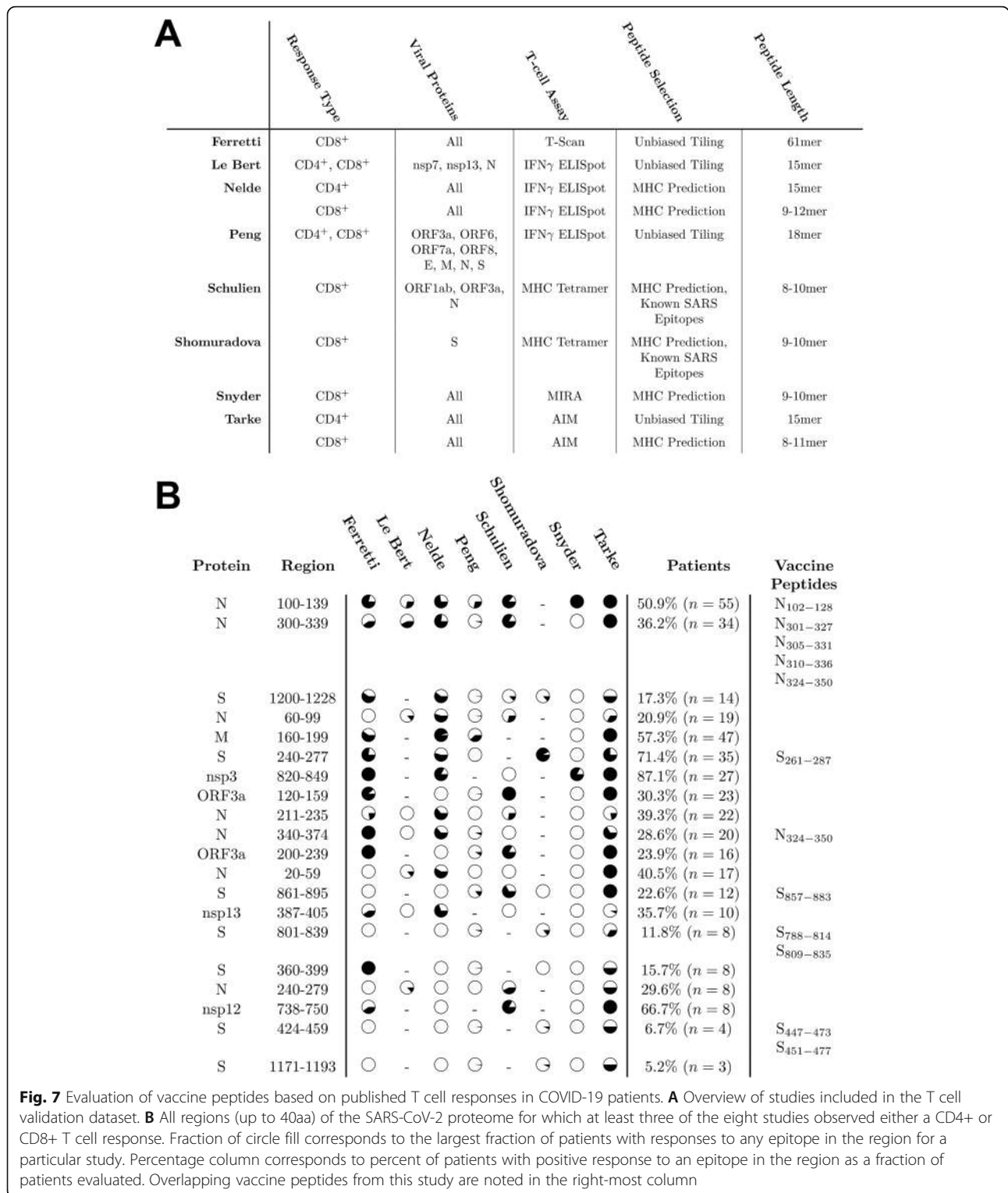## Murine validation of T and B cell epitope immunogenicity

We sought to experimentally evaluate our minimum set of predicted T and B cell epitope candidates. We manually selected 16 of the 22 vaccine peptides for synthesis, keeping at most 2 peptides per overlapping region with a preference for those with predicted H2$^d$ MHC ligands. We then vaccinated BALB/c mice with the 16 synthesized vaccine peptides and evaluated immune activation from humoral and T cell perspectives. Mice were vaccinated on experimental day 1, given booster vaccination on day 7, and sacrificed on day 21. We performed IFN-γ ELISpot in order to assess T cell activation by culturing splenocytes from vaccinated animals alongside each of the peptides within the vaccine pool. We observed a statistically significant increase in IFN-γ release in response to seven out of ten of our predicted T cell epitopes in mice vaccinated with peptides plus poly(I:C) versus poly(I:C) alone (Fig. 8A). We did not observe a statistically significant response against any of our six predicted B cell epitopes in our peptide vaccination group versus adjuvant alone. For evaluation of antibody responses, peptide (Fig. 8B) and S protein (Fig. 8C) ELISA from the day 21 sera of the above mice failed to show signal above adjuvant alone in all groups.

## Discussion

We report here a survey of the SARS-CoV-2 epitope landscape along with a strategy for prioritizing both T cell and B cell epitopes for vaccine development. Major vaccine efforts targeting coronaviruses have focused primarily on generation of neutralizing antibody responses [168–176]. CD4+ T cells provide help to B cells to support class switching, maturation, and antibody production. Additionally, they promote CD8+ T cell activation, maturation, and effector function. We therefore searched for vaccine peptide sequences which include both B cell epitopes and MHC ligands predicted to drive CD4+ and CD8+ T cell responses at high population frequencies within the U.S. based on data available in the first few months of the pandemic. Our current efforts are focused on testing the immunogenicity of these peptides in murine models, comparing those which contain overlapping and non-overlapping T and B cell epitopes. Results from such preclinical testing will inform an envisioned phase I clinical trial using a condensed peptide set targeting B cell epitopes with known viral neutralization plus optimal T cell epitopes.

Prior work has surveyed the epitope space of SARS-CoV-2 using analysis of sequence homology with SARS-CoV-1 epitopes, prediction of linear B cell epitopes, and prediction of T cell epitopes using IEDB tools. Grifoni et al. reported predicted T and B cell epitopes based on cross-referencing of known SARS epitopes with sequence homology to SARS-CoV-2 against SARS-CoV-2-specific parallel computational prediction [177]. This study did not consider epitope mapping of SARS-CoV-2 convalescent antibody repertoires, which may be

**Fig. 7** Evaluation of vaccine peptides based on published T cell responses in COVID-19 patients. **A** Overview of studies included in the T cell validation dataset. **B** All regions (up to 40aa) of the SARS-CoV-2 proteome for which at least three of the eight studies observed either a CD4+ or CD8+ T cell response. Fraction of circle fill corresponds to the largest fraction of patients with responses to any epitope in the region for a particular study. Percentage column corresponds to percent of patients with positive response to an epitope in the region as a fraction of patients evaluated. Overlapping vaccine peptides from this study are noted in the right-most column

important to achieve high specificity of B cell epitope predictions. Our prediction of T cell epitopes is conceptually similar to their computational process, but our study does not focus on conserved epitopes relative to SARS-CoV-1. Instead, we attempt to filter CD4+, CD8+,

and B cell epitopes by additional considerations of vaccination suitability (e.g., polymorphism, accessibility) and go beyond epitope selection to vaccine peptides integrating different categories of epitopes. Ahmed et al. reported a set of predicted T and B cell SARS-CoV-2
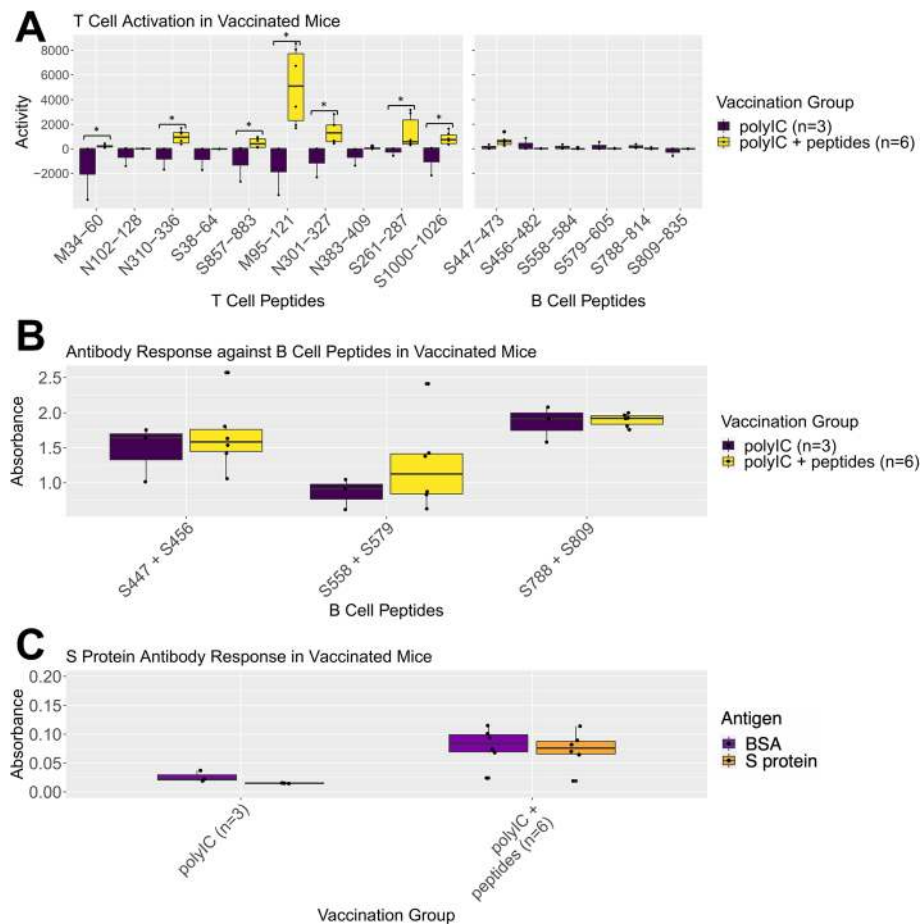
**Fig. 8** Experimental assessment of T and B cell epitope immunogenicity. **A** Mice were vaccinated with sixteen predicted T cell and B cell epitopes, designated as "peptides," in combination with poly(I:C), or with poly(I:C) alone. T cell activity in response to vaccination was measured via IFN-γ ELISpot with splenocytes isolated from mice at experimental day 21, plated with individual peptides. Activity was calculated by ELISpot plate reader. Peptide designations indicate protein, start, and end as shown in Fig. 6B. **B** Antibody response against predicted B cell peptide epitopes was measured via peptide ELISA. Wells were coated with pairs of predicted B cell peptides. **C** Antibody response against S protein was assessed via whole protein ELISA. Response to bovine serum albumin (BSA) was measured as negative control. For all subfigures, asterisks indicate statistically significant p value (< 0.05) from Mann-Whitney U tests of poly(I:C) + peptide groups compared to poly(I:C) alone

epitopes with associated assay confirmation within the NIAID ViPR database. However, these predicted epitopes were largely limited to those with sequence homology between SARS-CoV-1 and SARS-CoV-2, given the paucity of available SARS-CoV-2 assay data in the spring of 2020. Several studies identified linear B cell epitopes on the SARS-CoV-2 surface glycoprotein from sera of viral exposed patients using peptide arrays [55, 56, 58] as well as phage immunoprecipitation sequencing (PhIP-Seq) [57]. These studies are an important source of information, but it has also been shown that antibodies which recognize peptides often cross-react primarily with proteins only in denatured conformations [178–180]. There is a risk that identified linear epitopes would not be able to promote viral neutralization in vivo due to a lack of surface exposure. Our work adds to this important emerging field by analyzing the SARS-CoV-2

HLA ligand landscape through binding affinity filters derived from validated IEDB HLA ligands, as well as deriving T and B cell vaccine candidates through rational filtering criteria grounded in SARS-CoV-2 biology, including predicted immunogenicity, epitope location, glycosylation sites, and polymorphic sites. Additionally, inclusion of corresponding murine epitopes allows for future studies to be performed in animal models of SARS-CoV-2. We expect the application of these filters will improve specificity of antiviral response.

Other computational methods for prediction of SARS-CoV-2 epitopes have been described [181–183] in a continuously growing body of literature. Many of these studies consider population-specific MHC allele frequencies and attempt to derive an optimal epitope set that allows for broad population coverage. Liu et al. [182] adds to this by further considering allelic linkage

disequilibrium. Omnibus analysis of peptide-MHC binding from previously described tools was used to identify their peptide set comprising 19 each of MHC-I and MHC-II ligands. This method differs from our strategy in two ways: only considering peptide-MHC binding prediction rather than filtering for putative T cell epitopes, and deriving a set of 38 total minimal MHC-I and MHC-II ligands rather than identifying longer regions in the SARS-CoV-2 proteome that encompass regions with population-optimized T cell epitopes. While our capacity to predict peptide-MHC binding is reasonably accurate for MHC-I and variably accurate for MHC-II, our capacity to predict immunogenicity of any given minimal epitope remains limited. As such, we believe vaccinating with a longer (27mer) sequence containing multiple predicted minimal epitopes allows for a degree of purposeful imprecision, allowing for the optimal MHC-I and MHC-II sequences to be processed and presented in vivo. Compared to Poran et al. [181], which used a mass spectrometry-derived HLA presentation predictor, this peptide set is filtered through tetramer derived immunogenicity prediction—a more direct metric for epitope efficacy. Yarmarkovich et al. [183] addresses concerns for peptide immunogenicity versus auto-immunity by comparing predicted epitopes against a reference human peptidome.

While this study also filters for peptide overlap with self-epitopes, our immunogenicity prediction algorithm primarily considers peptide sequence features inspired by Calis et al. [146], predicted MHC scores, as well as the MHCflurry 2.0 [184] peptide processing score for CD8+ T cell epitopes, which are then used to fit a model against a validated viral tetramer dataset curated from IEDB [185]. Additionally, B cell epitopes were derived from in silico methods in Yarmarkovich et al., while this study used in vitro epitope mapping studies as the basis for our B cell epitope candidate set. Lastly, Gao et al. [186] approach the problem of SARS-CoV-2 epitope prediction by directly evaluating a candidate peptide's sequence similarity to both the human proteome and the set of pathogenic epitopes in IEDB; based on the methodology, Luksza et al. [187] used for cancer neoantigen prediction. This approach is intrinsically limited by a hypothetical sequence homology between T cell epitopes in SARS-CoV-2 and previously identified pathogenic epitopes. On the other hand, we use a diverse set of peptide-MHC features and do not expect actual sequence homology with any existing known epitopes.

A key aspect of our epitope selection process is the prioritization of overlapping CD4+, CD8+, and B cell epitopes. As the role of T cell epitope vaccines in SARS-CoV-2 continues to be investigated in model systems, we furthermore cross-referenced human and murine T cell epitopes to allow for murine vaccine studies using human-relevant peptides in H2b and H2d haplotypes. We hypothesize that inclusion of CD8+ epitopes may allow for clearance of SARS-CoV-2 from infected cells, and the inclusion of CD4+ epitopes may allow for greater activation of both cytotoxic and humoral antiviral responses. Overlapping CD4+ and CD8+ epitopes allowed for selection of peptide candidates covering a large proportion of the population. We next attempted to identify candidates with overlapping CD4+/CD8+ epitopes with B cell epitopes. However, these candidate options were limited due to the paucity of predicted B cell candidates. Therefore, a more effective strategy would be to include overlapping CD4+/CD8+ optimized peptides together with separate B cell optimized peptides. We expect this to provide the most robust and broad antiviral adaptive immune coverage by activating CD4+ T cells, CD8+ T cells, and B cells.

To this end, we predicted and tested the immunogenicity of peptides optimized for both overlapping CD4+/CD8+ T cell epitopes as well as peptides optimized for B cell epitopes. We observed statistically significant T cell activation measured by IFN-y release in response to seven of our 10 predicted T cell stimulatory epitopes when administered with poly(I:C) adjuvant as compared to vaccination with poly(I:C) alone. None of our six predicted B cell epitopes generated significant T cell activation, indicating that our method for predicting T cell immunogenicity is appropriately specific. A 70% success rate for prediction of T cell epitopes that would activate T cells to generate significantly enhanced IFN-y release demonstrates that our computational prediction of peptide vaccines was successful from a T cell standpoint. Further studies to assess (1) the CD4+ versus CD8+ responses against each peptide, (2) immunogenicity of individual epitopes within each peptide, and (3) the protective capacities of these epitopes are required to validate their therapeutic potential.

Contrasting these T cell findings, we did not observe increased antibody response against any of our predicted B cell epitopes in peptide-vaccinated mice compared to those vaccinated with adjuvant alone. We also did not observe any significant antibody response against S protein above negative control in vaccinated mice. This indicates that while our strategy was successful in predicting immunogenic T cell epitopes, our predicted B cell epitopes did not provide robust B cell activation by day 21. Options to further investigate these results include titrating dosage of the administered B cell peptides to evaluate whether concentrations used were sufficient to generate robust antibody responses, or further refinement of our criteria for B cell epitope prediction in order to predict epitopes more likely to generate an antibody response. Whether T cell responses in absence of antibody responses are sufficient for antiviral protection

remains unclear and can be addressed in future viral challenge studies.

In addition to epitope selection, optimal adjuvant choice for a SARS-CoV-2 vaccine is currently unclear. Prior evidence from SARS-CoV-1 suggested a Th2 dominant response to be associated with worse outcomes [13]—thus, adjuvant selection may also play an important role in SARS-CoV-2 in skewing the helper arm toward a Th1 phenotype. Patients with severe COVID-19 demonstrate elevated levels of CCR6+ Th17 cells [188]. Additionally, many COVID-19 patients with acute respiratory distress syndrome (ARDS) demonstrated cytokine storm manifested by elevation of a variety of cytokines, of which several are involved in Th17 responses [189]. In MERS patients, increased IL-17 to type I IFN is associated with worse outcome [190]. Altogether, the Th17 response may contribute to increased risk of severe pulmonary injury and worse outcomes in COVID-19 patients [191]. As the Th1 and Th17 cellular response pathways are closely linked, cotherapies that inhibit Th17 activation (e.g., secukinumab, tocilizumab) have been proposed for use in COVID-19; however, the efficacy of these therapies remains to be seen. The role of other helper subsets (Th9, Th18) remains even more poorly understood. Relevant for the vaccine studies presented here, poly(I:C) appears to primarily activate Th1 cells, skewing the immune response toward a phenotype that may be most beneficial [192]. Further studies would be needed to assess which subtypes of T cells were activated by our vaccine formulations.

One limitation of our study is that, while we use epitope mapping data with direct biological evidence for B cell epitopes in SARS-CoV-2, the T cell epitopes we report were all derived from computational prediction. In an effort to partially overcome this weakness, we applied binding affinity and immunogenicity prediction filters grounded in validated IEDB binding and tetramer studies. Other filtering criteria for T cell epitopes have been evaluated, including allergenicity, antigenicity, stability, and inflammatory/cytotoxic response [193–195]; it remains to be seen if these or other filtering criteria improve T cell epitope selection in SARS-CoV-2. Reassuringly, our selection of T cell-directed vaccine peptides demonstrates significant overlap with the recurrent epitopes identified in eight different studies examining T cell responses in COVID-19 patients (Fig. 7). Le Bert et al. looked for T cell epitopes within the nucleocapsid (N), nsp7 and nsp13 proteins in PBMCs of recovered COVID-19 patients using an IFN-γ ELISpot assay [196]. They identified two recurrent epitope regions (N101-120, N321-340) which overlap with multiple 27mer vaccine peptides in this paper (Fig. 6B, peptides 4–8). Shomuradova et al. also identified COVID-19

patient T cell epitopes, but using A*02:01 tetramers loaded with 13 distinct peptides from the surface glycoprotein (S) [71]. Two of these 13 peptides showed recurrent reactivity across 14 A*02:01-positive patients (S269-277 and S1000-1008). Both of these epitopes are also included in multiple 27mer vaccine peptides (Fig. 6B, peptides 11 and 15). Across all eight studies considered, the most recurrently identified epitopes fall within two regions, both in the nucleocapsid protein (N) around positions N100 and N300 (Fig. 7B), overlapping with multiple vaccine peptides selected by our algorithm. It is worth noting that our heuristic for selecting abundant proteins (only considering epitopes and vaccine peptides from the M, N, and S proteins) was moderately successful in that 15/20 recurrent epitope regions occurred in these proteins. While we missed recurrent epitope regions in ORF3a, nsp3, nsp12, and nsp13, filtering our predictions to the most abundant proteins allowed us to avoid many false positive predictions from ORF1ab and perform much better in predicting true T cell epitopes.

It is worth noting that the dataset of biologically measured T cell responses to SARS-CoV-2 infection which we curated to evaluate our vaccine peptide selection overlaps significantly with another study by Quadeer et al. [197]. The biggest difference between their approach and ours is that we do not require HLA restriction of identified epitope regions and can thus use a larger number of epitopes from assays such as unbiased ELISPOT screening. Since our evaluation seeks primarily to ascertain whether our vaccine peptides are highly enriched for immunogenic epitopes, we are less stringent in knowing exactly which epitopes are present and to which HLA alleles they bind.

A different potential limitation of this study is the insensitivity of our experiments to the total potential space of SARS-CoV-2 antibody epitopes. Our B cell epitope analyses start with only 58 identified linear antibody epitopes on the surface glycoprotein of SARS-CoV-2, while it is likely that many other epitopes are possible. Second, these linear epitope mappings do not allow for identification of antibodies which bind tertiary/quaternary protein structures. Lastly, identification of epitopes via array studies depended on differences in antibody binding to potential linear epitopes between uninfected and infected persons. There may be some cross-reactivity between antibodies generated against other coronaviruses and SARS-CoV-2, which if present might show reactivity in our screening assay. If true, our strategy would not identify these epitopes as specific for SARS-CoV-2. Similarly, we excluded viral regions with significant polymorphism across the viral population. We instead focused on conserved regions of SARS-CoV-2 to identify epitopes that would be most broadly targetable in the human population. For these reasons, we do not present

our antibody data as describing the complete set of SARS-CoV-2 epitopes.

## Conclusions

Our study sought to design a peptide vaccine for SARS-CoV-2 targeting immune responses from B cells, CD4$^+$ T cells, and CD8$^+$ T cells. This kind of vaccine may be a useful addition to the evolving landscape of SARS-CoV-2 vaccines since its rapid manufacturing and precise design may help fill gaps in immunity that arise due to antigenic drift of new viral variants. However, we emphasize that epitope selection is only one aspect of the problem, and a key question is whether a peptide vaccine can be sufficiently immunogenic. Adjuvant selection, conjugation to carriers such as KLH [43] or rTTHC [198], and prime/boost approaches using orthogonal platforms are all potential avenues to explore. Thus far, we have demonstrated the immunogenic capacity of our T cell epitope selection process coupled with linear peptide vaccination using poly(I:C) as an adjuvant. It is possible that the selected B cell epitopes in this work may still be useful for eliciting neutralizing responses when encoded using a more conformationally stable immunogen. We anticipate that the sets of vaccine peptides reported here may be valuable in the preclinical development of these approaches.

### Abbreviations
RBD: Receptor binding domain; ACE2: Angiotensin-converting enzyme 2; IEDB: Immune Epitope Database; PhIP-Seq: Phage immunoprecipitation sequencing; EL: Elution ligand; BA: Binding affinity; S: Surface glycoprotein; N: Nucleocapsid protein; M: Membrane; FP: Fusion peptide; HR: Heptad repeat; RBM: Receptor binding motif

### Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13073-021-00910-1.

---

**Additional file 1.** Contains all supplemental figures (Fig. S1 - S10).

**Additional file 2: Table S1.** All SARS-CoV-2 MHC-I ligands contained in the top 5% of U.S. HLA alleles. **Table S2.** All SARS-CoV-2 MHC-II ligands contained in the top 5% of U.S. HLA alleles. **Table S3.** All SARS-CoV-2 MHC-I ligands contained in the top 5% of worldwide HLA alleles. **Table S4.** All SARS-CoV-2 MHC-II ligands contained in the top 5% of worldwide HLA alleles. **Table S5.** Summary of SARS-CoV-1 MHC ligands previously described in the literature. **Table S6.** SARS-CoV-2 B cell linear epitopes from array/mapping studies. **Table S7.** SARS-CoV-2 T cell epitopes within S, M, and N proteins. **Table S8.** SARS-CoV-2 T cell epitopes within all proteins. **Table S9.** Curated a dataset of published T cell epitope mapping studies.

---

### Authors' contributions
C.C.S., K.S.O., K.M.G., S.E., C.W., S.V., J.K., T.O., J.W., B.G.V., and A.R. contributed to the conception and design of the work. C.C.S., K.S.O., K.M.G., M.S., W.B., J.G., S.E., C.W., S.V., A.W., M.F., B.C., E.R., J.K., T.O., C.H., K.H., V.S., E.G., A.M.S., J.P.T., H.W., O.C.G., R.J.W., K.K., M.H., B.G.V., and A.R. contributed to the acquisition, analysis, and interpretation of data. C.C.S., K.S.O., B.G.V., and A.R. have drafted the work or substantively revised it. The author(s) read and approved the final manuscript.

### Availability of data and materials
The datasets generated and/or analysed during the current study are available in the Vincent lab github repository, https://github.com/Benjamin-Vincent-Lab/Landscape-and-Selection-of-Vaccine-Epitopes-in-SARS-CoV-2 [199]. Several data files larger than 100 Mb and supplemental tables are available at https://data.mendeley.com/datasets/c6pdfrwxgj/6 [200].

## Declarations

### Ethics approval and consent to participate
The studies involving human participants were reviewed and approved by Ethics Committee of Charité Universitätsmedizin Berlin (EA2/066/20, EA1/068/20) and Ethics Committee at the Medical Faculty of the Ludwig Maximilians Universität Munich (vote 20-225 KB). The patients/participants provided their written informed consent to participate in this study. All murine experiments described in this study were approved by the UNC Institutional Animal Care and Use Committee (IACUC), ID 20-121.0. The research was conducted in strict compliance with the Helsinki Declaration.

### Consent for publication
Not applicable.

### Competing interests
C.H., K.H., and V.S. are employees of PEPperPRINT GmbH. The other authors declare that they have no competing interests.

### Author details
[1]Department of Microbiology and Immunology, UNC School of Medicine, Chapel Hill, NC, USA. [2]Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, CB# 7295, Chapel Hill, NC 27599-7295, USA. [3]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [4]PEPperPRINT GmbH, Heidelberg, Germany. [5]Genomics Institute, University of California, Santa Cruz, CA, USA. [6]Department of Genetics, UNC School of Medicine, Chapel Hill, NC, USA. [7]Institute for Inflammatory Diseases, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [8]Center for Translational Immunology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [9]Division of Medical Oncology, Department of Medicine, UNC School of Medicine, Chapel Hill, NC, USA. [10]Department of Biochemistry and Biophysics, UNC School of Medicine, Chapel Hill, NC, USA. [11]Complex Carbohydrate Research Center, University of Georgia, Athens, GA, USA. [12]Computational Medicine Program, UNC School of Medicine, Chapel Hill, NC, USA. [13]Curriculum in Bioinformatics and Computational Biology, UNC School of Medicine, Chapel Hill, NC, USA. [14]Division of Hematology, Department of Medicine, UNC School of Medicine, Chapel Hill, NC, USA.

### References
1. Graham BS. Advances in antiviral vaccine development. Immunol Rev. 2013; 255:230–42. https://doi.org/10.1111/imr.12098.
2. Hodgson J. The pandemic pipeline. Nat Biotechnol. 2020. https://doi.org/10.1038/d41587-020-00005-z.
3. With record-setting speed, vaccinemakers take their first shots at the new coronavirus. Science | AAAS. 2020. https://www.sciencemag.org/news/2020/03/record-setting-speed-vaccine-makers-take-their-first-shots-new-coronavirus. Accessed 3 Apr 2020.
4. Tai W, He L, Zhang X, Pu J, Voronin D, Jiang S, et al. Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for

Smith *et al. Genome Medicine*        (2021) 13:101

Page 19 of 23

development of RBD protein as a viral attachment inhibitor and vaccine. Cell Mol Immunol. 2020. https://doi.org/10.1038/s41423-020-0400-4.

5.  Piccoli L, Park Y-J, Tortorici MA, Czudnochowski N, Walls AC, Beltramello M, et al. Mapping neutralizing and immunodominant sites on the SARS-CoV-2 spike receptor-binding domain by structure-guided high-resolution serology. Cell. 2020;183:1024–42.e21. https://doi.org/10.1016/j.cell.2020.09.037.

6.  Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. Nature. 2020;584: 450–6. https://doi.org/10.1038/s41586-020-2571-7.

7.  Laczkó D, Hogan MJ, Toulmin SA, Hicks P, Lederer K, Gaudette BT, et al. A single immunization with nucleoside-modified mRNA vaccines elicits strong cellular and humoral immune responses against SARS-CoV-2 in mice. Immunity. 2020. https://doi.org/10.1016/j.immuni.2020.07.019.

8.  Zang J, Gu C, Zhou B, Zhang C, Yang Y, Xu S, et al. Immunization with the receptor-binding domain of SARS-CoV-2 elicits antibodies cross-neutralizing SARS-CoV-2 and SARS-CoV without antibody-dependent enhancement. Cell Discov. 2020;6:61. https://doi.org/10.1038/s41421-020-00199-1.

9.  Corbett KS, Flynn B, Foulds KE, Francica JR, Boyoglu-Barnum S, Werner AP, et al. Evaluation of the mRNA-1273 vaccine against SARS-CoV-2 in nonhuman primates. N Engl J Med. 2020;383:1544–55. https://doi.org/10.1056/NEJMoa2024671.

10.  Jiang S, Hillyer C, Du L. Neutralizing antibodies against SARS-CoV-2 and other human coronaviruses. Trends Immunol. 2020;41:355–9. https://doi.org/10.1016/j.it.2020.03.007.

11.  Swain SL, McKinstry KK, Strutt TM. Expanding roles for CD4+ T cells in immunity to viruses. Nat Rev Immunol. 2012;12:136–48. https://doi.org/10.1038/nri3152.

12.  Kulinski JM, Tarakanova VL, Verbsky J. Regulation of antiviral CD8 T-cell responses. Crit Rev Immunol. 2013;33:477–88. https://doi.org/10.1615/critrevimmunol.2013007909.

13.  Li CK-F, Wu H, Yan H, Ma S, Wang L, Zhang M, et al. T cell responses to whole SARS coronavirus in humans. J Immunol. 2008;181:5490–500. https://doi.org/10.4049/jimmunol.181.8.5490.

14.  Jeyanathan M, Afkhami S, Smaill F, Miller MS, Lichty BD, Xing Z. Immunological considerations for COVID-19 vaccine strategies. Nat Rev Immunol. 2020;20:615–32. https://doi.org/10.1038/s41577-020-00434-6.

15.  Spellberg B, Edwards JE Jr. Type 1/Type 2 immunity in infectious diseases. Clin Infect Dis. 2001;32:76–102. https://doi.org/10.1086/317537.

16.  Gil-Etayo FJ, Suàrez-Fernández P, Cabrera-Marante O, Arroyo D, Garcinuño S, Naranjo L, et al. T-helper cell subset response is a determining factor in COVID-19 progression. Front Cell Infect Microbiol. 2021;11:624483. https://doi.org/10.3389/fcimb.2021.624483.

17.  Thanh Le T, Andreadakis Z, Kumar A, Gómez Román R, Tollefsen S, Saville M, et al. The COVID-19 vaccine development landscape. Nat Rev Drug Discov. 2020. doi:https://doi.org/10.1038/d41573-020-00073-5.

18.  van Doremalen N, Lambe T, Spencer A, Belij-Rammerstorfer S, Purushotham JN, Port JR, et al. ChAdOx1 nCoV-19 vaccination prevents SARS-CoV-2 pneumonia in rhesus macaques. bioRxiv. 2020:2020.05.13.093195. https://doi.org/10.1101/2020.05.13.093195.

19.  Yu J, Tostanoski LH, Peter L, Mercado NB, McMahan K, Mahrokhian SH, et al. DNA vaccine protection against SARS-CoV-2 in rhesus macaques. Science. 2020. https://doi.org/10.1126/science.abc6284.

20.  Zhu F-C, Li Y-H, Guan X-H, Hou L-H, Wang W-J, Li J-X, et al. Safety, tolerability, and immunogenicity of a recombinant adenovirus type-5 vectored COVID-19 vaccine: a dose-escalation, open-label, non-randomised, first-in-human trial. Lancet. 2020. https://doi.org/10.1016/s0140-6736(20)31208-3.

21.  Keech C, Albert G, Cho I, Robertson A, Reed P, Neal S, et al. Phase 1–2 trial of a SARS-CoV-2 recombinant spike protein nanoparticle vaccine. N Engl J Med. 2020;383:2320–32. https://doi.org/10.1056/NEJMoa2026920.

22.  Baden LR, El Sahly HM, Essink B, Kotloff K, Frey S, Novak R, et al. Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine. N Engl J Med. 2021;384:403–16. https://doi.org/10.1056/NEJMoa2035389.

23.  Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine. N Engl J Med. 2020;383:2603–15. https://doi.org/10.1056/NEJMoa2034577.

24.  Voysey M, Clemens SAC, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Safety and efficacy of the ChAdOx1 nCoV-19 vaccine (AZD1222) against SARS-CoV-2: an interim analysis of four randomised controlled trials in Brazil, South Africa, and the UK. Lancet. 2021;397:99–111. https://doi.org/10.1016/S0140-6736(20)32661-1.

25.  Ella R, Vadrevu KM, Jogdand H, Prasad S, Reddy S, Sarangi V, et al. Safety and immunogenicity of an inactivated SARS-CoV-2 vaccine, BBV152: a double-blind, randomised, phase 1 trial. Lancet Infect Dis. 2021. https://doi.org/10.1016/S1473-3099(20)30942-7.

26.  Barnes CO, West AP Jr, Huey-Tubman KE, Hoffmann MAG, Sharaf NG, Hoffman PR, et al. Structures of human antibodies bound to SARS-CoV-2 spike reveal common epitopes and recurrent features of antibodies. Cell. 2020;182:828–42.e16. https://doi.org/10.1016/j.cell.2020.06.025.

27.  Logunov DY, Dolzhikova IV, Shcheblyakov DV, Tukhvatulin AI, Zubkova OV, Dzharullaeva AS, et al. Safety and efficacy of an rAd26 and rAd5 vector-based heterologous prime-boost COVID-19 vaccine: an interim analysis of a randomised controlled phase 3 trial in Russia. Lancet. 2021;397:671–81. https://doi.org/10.1016/S0140-6736(21)00234-8.

28.  Voysey M, Costa Clemens SA, Madhi SA, Weckx LY, Folegatti PM, Aley PK, et al. Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of ChAdOx1 nCoV-19 (AZD1222) vaccine: a pooled analysis of four randomised trials. Lancet. 2021; 397:881–91. https://doi.org/10.1016/S0140-6736(21)00432-3.

29.  Gribble J, Stevens LJ, Agostini ML, Anderson-Daniels J, Chappell JD, Lu X, et al. The coronavirus proofreading exoribonuclease mediates extensive viral recombination. Plos Pathog. 2021;17:e1009226. https://doi.org/10.1371/journal.ppat.1009226.

30.  Clark SA, Clark LE, Pan J, Coscia A, McKay LGA, Shankar S, et al. SARS-CoV-2 evolution in an immunocompromised host reveals shared neutralization escape mechanisms. Cell. 2021. https://doi.org/10.1016/j.cell.2021.03.027.

31.  Edara VV, Norwood C, Floyd K, Lai L, Davis-Gardner ME, Hudson WH, et al. Infection- and vaccine-induced antibody binding and neutralization of the B.1.351 SARS-CoV-2 variant. Cell Host Microbe. 2021. https://doi.org/10.1016/j.chom.2021.03.009.

32.  Cele S, Gazy I, Jackson L, Hwa S-H, Tegally H, Lustig G, et al. Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. Nature. 2021. https://doi.org/10.1038/s41586-021-03471-w.

33.  Hoffmann M, Arora P, Groß R, Seidel A, Hörnich BF, Hahn AS, et al. SARS-CoV-2 variants B.1.351 and P.1 escape from neutralizing antibodies. Cell. 2021. https://doi.org/10.1016/j.cell.2021.03.036.

34.  Tarke A, Sidney J, Methot N, Zhang Y, Dan JM, Goodwin B, et al. Negligible impact of SARS-CoV-2 variants on CD4+ and CD8+ T cell reactivity in COVID-19 exposed donors and vaccinees. bioRxiv. 2021;2021.02.27.433180. doi:https://doi.org/10.1101/2021.02.27.433180.

35.  Teijaro JR, Farber DL. COVID-19 vaccines: modes of immune activation and future challenges. Nat Rev Immunol. 2021;21:195–7. https://doi.org/10.1038/s41577-021-00526-x.

36.  Sherina N, Piralla A, Du L, Wan H, Kumagai-Braesch M, Andréll J, et al. Persistence of SARS-CoV-2-specific B and T cell responses in convalescent COVID-19 patients 6–8 months after the infection. Med. 2021;2:281–95.e4. https://doi.org/10.1016/j.medj.2021.02.001.

37.  Khong H, Volmari A, Sharma M, Dai Z, Imo CS, Hailemichael Y, et al. Peptide vaccine formulation controls the duration of antigen presentation and magnitude of tumor-specific CD8+ T cell response. J Immunol. 2018;200: 3464–74. https://doi.org/10.4049/jimmunol.1700467.

38.  Baz A, Buttigieg K, Zeng W, Rizkalla M, Jackson DC, Groves P, et al. Branched and linear lipopeptide vaccines have different effects on primary CD4 and CD8 T-cell activation but induce similar tumor-protective memory CD8 T-cell responses. Vaccine. 2008;26:2570–9. https://doi.org/10.1016/j.vaccine.2008.03.022.

39.  Martins KAO, Cooper CL, Stronsky SM, Norris SLW, Kwilas SA, Steffens JT, et al. Adjuvant-enhanced CD4 T cell responses are critical to durable vaccine immunity. EBioMedicine. 2016;3:67–78. https://doi.org/10.1016/j.ebiom.2015.11.041.

40.  Rosalia RA, Quakkelaar ED, Redeker A, Khan S, Camps M, Drijfhout JW, et al. Dendritic cells process synthetic long peptides better than whole protein, improving antigen presentation and T-cell activation. Eur J Immunol. 2013; 43:2554–65. https://doi.org/10.1002/eji.201343324.

41.  Wang CY, Chang TY, Walfield AM, Ye J, Shen M, Chen SP, et al. Effective synthetic peptide vaccine for foot-and-mouth disease in swine. Vaccine. 2002;20:2603–10. https://doi.org/10.1016/s0264-410x(02)00148-2.

42.  Zhou M, Kostoula I, Brill B, Panou E, Sakarellos-Daitsiotis M, Dietrich U. Prime boost vaccination approaches with different conjugates of a new HIV-1 gp41 epitope encompassing the membrane proximal external region induce neutralizing antibodies in mice. Vaccine. 2012;30:1911–6. https://doi.org/10.1016/j.vaccine.2012.01.026.

43. Langeveld JP, Casal JI, Osterhaus AD, Cortés E, de Swart R, Vela C, et al. First peptide vaccine providing protection against viral infection in the target animal: studies of canine parvovirus in dogs. J Virol. 1994;68:4506–13 https://www.ncbi.nlm.nih.gov/pubmed/8207825.

44. Vázquez S, Guzmán MG, Guillen G, Chinea G, Pérez AB, Pupo M, et al. Immune response to synthetic peptides of dengue prM protein. Vaccine. 2002;20:1823–30. https://doi.org/10.1016/s0264-410x(01)00515-1.

45. Vieillard V, Combadière B, Tubiana R, Launay O, Pialoux G, Cotte L, et al. HIV therapeutic vaccine enhances non-exhausted CD4+ T cells in a randomised phase 2 trial. NPJ Vaccines. 2019;4:25. https://doi.org/10.1038/s41541-019-0117-5.

46. Pavlick AC, Blazquez A, Meseck M, Donovan MJ, Castillo-Martin M, Htwe Thin T, et al. A phase II open labeled, randomized study of poly-ICLC matured dendritic cells for NY-ESO-1 and Mean-A peptide vaccination compared to Montanide, in melanoma patients in complete clinical remission. J Clin Orthod. 2019;37:9538. https://doi.org/10.1200/JCO.2019.37.15_suppl.9538.

47. Firbas C, Jilma B, Tauber E, Buerger V, Jelovcan S, Lingnau K, et al. Immunogenicity and safety of a novel therapeutic hepatitis C virus (HCV) peptide vaccine: a randomized, placebo controlled trial for dose optimization in 128 healthy subjects. Vaccine. 2006;24:4343–53. https://doi.org/10.1016/j.vaccine.2006.03.009.

48. Francis JN, Bunce CJ, Horlock C, Watson JM, Warrington SJ, Georges B, et al. A novel peptide-based pan-influenza A vaccine: a double blind, randomised clinical trial of immunogenicity and safety. Vaccine. 2015;33:396–402. https://doi.org/10.1016/j.vaccine.2014.06.006.

49. Pennington MW, Zell B, Bai CJ. Commercial manufacturing of current good manufacturing practice peptides spanning the gamut from neoantigen to commercial large-scale products. Med Drug Discov. 2021;9:100071. https://doi.org/10.1016/j.medidd.2020.100071.

50. Bray BL. Large-scale manufacture of peptide therapeutics by chemical synthesis. Nat Rev Drug Discov. 2003;2:587–93. https://doi.org/10.1038/nrd1133.

51. Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, et al. The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. 2019;47: D339–43. https://doi.org/10.1093/nar/gky1006.

52. Tian J-H, Patel N, Haupt R, Zhou H, Weston S, Hammond H, et al. SARS-CoV-2 spike glycoprotein vaccine candidate NVX-CoV2373 elicits immunogenicity in baboons and protection in mice. Cold Spring Harbor Lab. 2020;2020.06.29.178509. doi:https://doi.org/10.1101/2020.06.29.178509.

53. Smith TRF, Patel A, Ramos S, Elwood D, Zhu X, Yan J, et al. Immunogenicity of a DNA vaccine candidate for COVID-19. Nat Commun. 2020;11:2601. https://doi.org/10.1038/s41467-020-16505-0.

54. Dinnon KH 3rd, Leist SR, Schäfer A, Edwards CE, Martinez DR, Montgomery SA, et al. A mouse-adapted model of SARS-CoV-2 to test COVID-19 countermeasures. Nature. 2020;586:560–6. https://doi.org/10.1038/s41586-020-2708-8.

55. Heidepriem J, Dahlke C, Kobbe R, Santer R, Koch T, Fathi A, et al. Longitudinal Development of antibody responses in COVID-19 patients of different severity with ELISA, peptide, and glycan arrays: an immunological case series. Pathogens. 2021;10. https://doi.org/10.3390/pathogens10040438.

56. Wang H, Wu X, Zhang X, Hou X, Liang T, Wang D, et al. SARS-CoV-2 proteome microarray for mapping COVID-19 antibody interactions at amino acid resolution. ACS Cent Sci. 2020;6:2238–49. https://doi.org/10.1021/acscentsci.0c00742.

57. Zamecnik CR, Rajan JV, Yamauchi KA, Mann SA, Loudermilk RP, Sowa GM, et al. ReScan, a multiplex diagnostic pipeline, pans human sera for SARS-CoV-2 antigens. Cell Rep Med. 2020;1:100123. https://doi.org/10.1016/j.xcrm.2020.100123.

58. Poh CM, Carissimo G, Wang B, Amrun SN, Lee CY-P, Chee RS-L, et al. Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. Nat Commun. 2020;11:2806. https://doi.org/10.1038/s41467-020-16638-2.

59. Schwarz T, Heiss K, Mahendran Y, Casilag F, Kurth F, Sander LE, et al. SARS-CoV-2 proteome-wide analysis revealed significant epitope signatures in COVID-19 patients. Front Immunol. 2021;12:629185. https://doi.org/10.3389/fimmu.2021.629185.

60. Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/hum - Nucleotide - NCBI. https://www.ncbi.nlm.nih.gov/nuccore/MT072688. Accessed 1 Apr 2021.

61. Schaid DJ. HaploStats. Rochester: Mayo Clinic/Foundation; 2005.

62. Klitz W, Maiers M, Spellman S, Baxter-Lowe LA, Schmeckpeper B, Williams TM, et al. New HLA haplotype frequency reference standards: high-resolution and large sample typing of HLA DR-DQ haplotypes in a sample of European Americans. Tissue Antigens. 2003;62:296–307. https://doi.org/10.1034/j.1399-0039.2003.00103.x.

63. Jurtz V, Paul S, Andreatta M, Marcatili P. NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. The Journal of. 2017. https://www.jimmunol.org/content/199/9/3360.abstract. Accessed 21 May 2020.

64. O'Donnell TJ, Rubinsteyn A, Bonsack M, Riemer AB, Laserson U, Hammerbacher J. MHCflurry: open-source class I MHC binding affinity prediction. Cell Syst. 2018;7:129–32.e4. https://doi.org/10.1016/j.cels.2018.05.014.

65. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics. 2015;67:641–50. https://doi.org/10.1007/s00251-015-0873-y.

66. Reynisson B, Barra C, Kaabinejadian S, Hildebrand WH, Peters B, Nielsen M. Improved prediction of MHC II antigen presentation through integration and motif deconvolution of mass spectrometry MHC eluted ligand data. J Proteome Res. 2020. https://doi.org/10.1021/acs.jproteome.9b00874.

67. Zhou T, Wang H, Luo D, Rowe T, Wang Z, Hogan RJ, et al. An exposed domain in the severe acute respiratory syndrome coronavirus spike protein induces neutralizing antibodies. J Virol. 2004;78:7217–26. https://doi.org/10.1128/JVI.78.13.7217-7226.2004.

68. Tarke A, Sidney J, Kidd CK, Dan JM, Ramirez SI, Yu ED, et al. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. bioRxiv. 2020. https://doi.org/10.1101/2020.12.08.416750.

69. Schulien I, Kemming J, Oberhardt V, Wild K, Seidel LM, Killmer S, et al. Characterization of pre-existing and induced SARS-CoV-2-specific CD8+ T cells. Nat Med. 2020. https://doi.org/10.1038/s41591-020-01143-2.

70. Snyder TM, Gittelman RM, Klinger M, May DH, Osborne EJ, Taniguchi R, et al. Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. medRxiv. 2020. https://doi.org/10.1101/2020.07.31.20165647.

71. Shomuradova AS, Vagida MS, Sheetikov SA, Zornikova KV, Kiryukhin D, Titov A, et al. SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T-cell receptors. medRxiv. 2020; https://www.medrxiv.org/content/10.1101/2020.05.20.20107813v1.abstract. Accessed 8 May 2021.

72. Peng Y, Mentzer AJ, Liu G, Yao X, Yin Z, Dong D, 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. 2020;21:1336–45. https://doi.org/10.1038/s41590-020-0782-6.

73. Nelde A, Bilich T, Heitmann JS, Maringer Y, Salih HR, Roerden M, et al. SARS-CoV-2 T-cell epitopes define heterologous and COVID-19-induced T-cell recognition. 2020. https://www.researchsquare.com/article/rs-35331/latest.

74. Le Bert N, Tan AT, Kunasegaran K, Tham CYL, Hafezi M, Chia A, et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. Nature. 2020;584:457–62. https://doi.org/10.1038/s41586-020-2550-z.

75. Ferretti AP, Kula T, Wang Y, Nguyen DMV, Weinheimer A, Dunlap GS, et al. COVID-19 patients form memory CD8+ T cells that recognize a small set of shared immunodominant epitopes in SARS-CoV-2; 2020. https://doi.org/10.2139/ssrn.3669387.

76. Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. J Mol Biol. 1982;157:105–32 https://www.ncbi.nlm.nih.gov/pubmed/7108955.

77. Kyi C, Sabado RL, Blazquez A, Posner MR, Genden EM, Miles BA, et al. A phase I study of the safety and immunogenicity of a multipeptide personalized genomic vaccine in the adjuvant treatment of solid cancers. J Clin Orthod. 2017;35:TPS3114. https://doi.org/10.1200/JCO.2017.35.15_suppl.TPS3114.

78. Peptide Design Guideline. Biomatik; 2011. https://www.biomatik.com/content/service_docs/peptide_design_guideline.pdf. Accessed 9 Apr 2021.

79. Peptide design guidelines. SB Peptide. https://www.sb-peptide.com/support/design/. Accessed 9 Apr 2021.

80. Grant GA. Synthetic peptides: a user's guide: Oxford University Press; 2002. https://play.google.com/store/books/details?id=_Z3mCwAAQBAJ. Accessed 9 Apr 2021.

81. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017;1:33–46. https://doi.org/10.1002/gch2.1018.

Smith *et al. Genome Medicine*        (2021) 13:101

Page 21 of 23

82. Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34: 4121–3. https://doi.org/10.1093/bioinformatics/bty407.

83. Katoh K, Misawa K, Kuma K-I, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 2002;30:3059–66. https://doi.org/10.1093/nar/gkf436.

84. Bastola A, Sah R, Rodriguez-Morales AJ, Lal BK, Jha R, Ojha HC, et al. The first 2019 novel coronavirus case in Nepal. Lancet Infect Dis. 2020;20:279–80. https://doi.org/10.1016/S1473-3099(20)30067-0.

85. Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, et al. GenBank. Nucleic Acids Res. 2018;46:D41–7. https://doi.org/10.1093/nar/gkx1094.

86. Charif D, Lobry JR. SeqinR 1.0-2: A contributed package to the R Project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. Structural approaches to sequence evolution: molecules, networks, populations. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. p. 207–32. https://doi.org/10.1007/978-3-540-35306-5_10.

87. Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. msa: an R package for multiple sequence alignment. Bioinformatics. 2015;31:3997–9. https://doi.org/10.1093/bioinformatics/btv494.

88. Kuhn M. Classification and regression training [R package caret version 6.0-86]. https://CRAN.R-project.org/package=caret. Accessed 21 May 2020.

89. Wilke CO. cowplot: streamlined plot theme and plot annotations for "ggplot2." CRAN Repos; 2016.

90. Dowle M, Srinivasan A. data. table: Extension of "data. frame". R package version 1.10. 4-3. 2017. https://CRAN.R-project.org/package=data.table. Accessed 21 May 2020.

91. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.

92. Revolution Analytics WS. doMC: foreach parallel adaptor for "parallel". R package version 1.3. 4. 2015. https://CRAN.R-project.org/package=doMC. Accessed 21 May 2020.

93. Wickham H, Francois R, Henry L, Müller K. dplyr: a grammar of data manipulation. R package version 0.4. 3. R Found Stat Comput , Vienna https://CRAN R-project org/package= dplyr. 2015. https://CRAN.R-project.org/package=dplyr. Accessed 21 May 2020.

94. Wickham H. forcats: tools for working with categorical variables (factors). 2017. https://CRANR-projectorg/package=forcatsRpackageversion020. Accessed 21 May 2020.

95. Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, et al. Software for computing and annotating genomic ranges. Plos Comput Biol. 2013;9:e1003118. https://doi.org/10.1371/journal.pcbi.1003118.

96. Pav SE. Grab Bag of "ggplot2" Functions [R package ggallin version 0.1.1]. https://CRAN.R-project.org/package=ggallin. Accessed 21 May 2020.

97. Clarke E, Sherrill-Mix S. Ggbeeswarm: categorical scatter (violin point) plots. R package version 0 6 0 Retrieved from https://CRANR-projectorg. 2017. Accessed 21 May 2020.

98. Campitelli E. Multiple fill and colour scales in "ggplot2" [R package ggnewscale version 0.4.1]. https://CRAN.R-project.org/package=ggnewscale. Accessed 21 May 2020.

99. Kassambara A. "ggplot2" based publication ready plots [R package ggpubr version 0.3.0]. https://CRAN.R-project.org/package=ggpubr. Accessed 21 May 2020.

100. Slowikowski K. Automatically position non-overlapping text labels with "ggplot2" [R package ggrepel version 0.8.2]. https://CRAN.R-project.org/package=ggrepel. Accessed 21 May 2020.

101. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, et al. gplots: Various R programming tools for plotting data. 2015. https://www.scienceopen.com/document?vid=0e5d8e31-1fe4-492f-a3d8-8cd71b2b8ad9. Accessed 21 May 2020.

102. Auguie B. Miscellaneous functions for "grid" graphics [R package gridExtra version 2.3]. https://CRAN.R-project.org/package=gridExtra. Accessed 21 May 2020.

103. Hugh-Jones D. HuxTable: Easily create and style tables for LaTeX, HTML and other formats. R package version 4.7. 1; 2019.

104. Bache SM, Wickham H. magrittr: a forward-pipe operator for R. R package version; 2014. p. 1.

105. Gohel D. officer: manipulation of Microsoft Word and PowerPoint documents; 2018.

106. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77. https://doi.org/10.1186/1471-2105-12-77.

107. Neuwirth E. RColorBrewer: ColorBrewer palettes. R package version 1.1-2. The R Foundation. 2014. https://CRAN.R-project.org/package=RColorBrewer. Accessed 21 May 2020.

108. Wickham H, Bryan J. readxl: Read excel files. R package version; 2019. p. 1.

109. Wickham H. scales: scale functions for visualization. R package version 0.4. 0. 2016. https://CRAN.R-project.org/package=scales. Accessed 21 May 2020.

110. Wickham H. stringr: Simple, consistent wrappers for common string operations (Package Version 1.2. 0)[Computer software]. 2017. https://CRAN.R-project.org/package=stringr. Accessed 21 May 2020.

111. CRAN - Package venneuler. https://CRAN.R-project.org/package=venneuler. Accessed 21 May 2020.

112. Garnier S. viridis: Default Color Maps from "matplotlib". 2016. R package version 0.3. 4. 2017. https://CRAN.R-project.org/package=viridis. Accessed 21 May 2020.

113. van der Walt S, Colbert SC, Varoquaux G. The NumPy Array: A Structure for Efficient Numerical Computation. Comput Sci Eng. 2011;13:22–30. https://doi.org/10.1109/MCSE.2011.37.

114. McKinney W, Others. pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing. 2011; 14. https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf. Accessed 21 May 2020.

115. Hunter JD. Matplotlib: a 2D graphics environment. Comput Sci Eng. 2007;9: 90–5. https://doi.org/10.1109/MCSE.2007.55.

116. Kluyver T, Ragan-Kelley B, Pérez F, Granger BE, Bussonnier M, Frederic J, et al. Jupyter Notebooks-a publishing format for reproducible computational workflows. In: ELPUB; 2016. p. 87–90. https://books.google.com/books?hl=en&lr=&id=Lgy3DAAAQBAJ&oi=fnd&pg=PA87&dq=jupyter&ots=N1GS7UpCdp&sig=YPTK4iGHEE712ql7APEF_IRSJBk. Accessed 21 May 2020.

117. Gonzalez-Galarza FF, McCabe A, Santos EJMD, Jones J, Takeshita L, Ortega-Rivera ND, et al. Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. 2020;48:D783–8. https://doi.org/10.1093/nar/gkz1029.

118. Sette A, Vitiello A, Reherman B, Fowler P, Nayersina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. J Immunol. 1994;153:5586–92 https://www.ncbi.nlm.nih.gov/pubmed/7527444.

119. Rajasagi M, Shukla SA, Fritsch EF, Keskin DB, DeLuca D, Carmona E, et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. Blood. 2014;124:453–62. https://doi.org/10.1182/blood-2014-04-567933.

120. Ahmed SF, Quadeer AA, McKay MR. Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses. 2020;12. https://doi.org/10.3390/v12030254.

121. Liu J, Sun Y, Qi J, Chu F, Wu H, Gao F, et al. The membrane protein of severe acute respiratory syndrome coronavirus acts as a dominant immunogen revealed by a clustering region of novel functionally and structurally defined cytotoxic T-lymphocyte epitopes. J Infect Dis. 2010;202: 1171–80. https://doi.org/10.1086/656315.

122. Liu J, Wu P, Gao F, Qi J, Kawana-Tachikawa A, Xie J, et al. Novel immunodominant peptide presentation strategy: a featured HLA-A*2402-restricted cytotoxic T-lymphocyte epitope stabilized by intrachain hydrogen bonds from severe acute respiratory syndrome coronavirus nucleocapsid protein. J Virol. 2010;84:11849–57. https://doi.org/10.1128/JVI.01464-10.

123. Ng O-W, Chia A, Tan AT, Jadi RS, Leong HN, Bertoletti A, et al. Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. Vaccine. 2016;34:2008–14. https://doi.org/10.1016/j.vaccine.2016.02.063.

124. Oh H-LJ, Chia A, Chang CXL, Leong HN, Ling KL, Grotenbreg GM, et al. Engineering T cells specific for a dominant severe acute respiratory syndrome coronavirus CD8 T cell epitope. J Virol. 2011;85:10464–71. https://doi.org/10.1128/JVI.05039-11.

125. Cheung Y-K, Cheng SC-S, Sin FW-Y, Chan K-T, Xie Y. Induction of T-cell response by a DNA vaccine encoding a novel HLA-A*0201 severe acute respiratory syndrome coronavirus epitope. Vaccine. 2007;25:6070–7. https://doi.org/10.1016/j.vaccine.2007.05.025.

126. Ohno S, Kohyama S, Taneichi M, Moriya O, Hayashi H, Oda H, et al. Synthetic peptides coupled to the surface of liposomes effectively induce SARS coronavirus-specific cytotoxic T lymphocytes and viral clearance in HLA-A*0201 transgenic mice. Vaccine. 2009;27:3912–20. https://doi.org/10.1016/j.vaccine.2009.04.001.

127. Røder G, Kristensen O, Kastrup JS, Buus S, Gajhede M. Structure of a SARS coronavirus-derived peptide bound to the human major histocompatibility

complex class I molecule HLA-B*1501. Acta Crystallogr Sect F Struct Biol Cryst Commun. 2008;64(Pt 6):459–62. https://doi.org/10.1107/S1744309108012396.

128. Du L, Zhao G, Lin Y, Chan C, He Y, Jiang S, et al. Priming with rAAV encoding RBD of SARS-CoV S protein and boosting with RBD-specific peptides for T cell epitopes elevated humoral and cellular immune responses against SARS-CoV infection. Vaccine. 2008;26:1644–51. https://doi.org/10.1016/j.vaccine.2008.01.025.

129. Tsao Y-P, Lin J-Y, Jan J-T, Leng C-H, Chu C-C, Yang Y-C, et al. HLA-A*0201 T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus nucleocapsid and spike proteins. Biochem Biophys Res Commun. 2006;344: 63–71. https://doi.org/10.1016/j.bbrc.2006.03.152.

130. Lv Y, Ruan Z, Wang L, Ni B, Wu Y. Identification of a novel conserved HLA-A*0201-restricted epitope from the spike protein of SARS-CoV. BMC Immunol. 2009;10:61. https://doi.org/10.1186/1471-2172-10-61.

131. Wang B, Chen H, Jiang X, Zhang M, Wan T, Li N, et al. Identification of an HLA-A*0201-restricted CD8+ T-cell epitope SSp-1 of SARS-CoV spike protein. Blood. 2004;104:200–6. https://doi.org/10.1182/blood-2003-11-4072.

132. Wang Y-D, Sin W-YF XG-B, Yang H-H, Wong T-Y, Pang X-W, et al. T-cell epitopes in severe acute respiratory syndrome (SARS) coronavirus spike protein elicit a specific T-cell immune response in patients who recover from SARS. J Virol. 2004;78:5612–8. https://doi.org/10.1128/JVI.78.11.5612-5618.2004.

133. Li T, Xie J, He Y, Fan H, Baril L, Qiu Z, et al. Long-term persistence of robust antibody and cytotoxic T cell responses in recovered patients infected with SARS coronavirus. Plos One. 2006;1:e24. https://doi.org/10.1371/journal.pone.0000024.

134. Chang CXL, Tan AT, Or MY, Toh KY, Lim PY, Chia ASE, et al. Conditional ligands for Asian HLA variants facilitate the definition of CD8+ T-cell responses in acute and chronic viral diseases. Eur J Immunol. 2013;43:1109–20. https://doi.org/10.1002/eji.201243088.

135. Chen H, Hou J, Jiang X, Ma S, Meng M, Wang B, et al. Response of memory CD8+ T cells to severe acute respiratory syndrome (SARS) coronavirus in recovered SARS patients and healthy individuals. J Immunol. 2005;175:591–8. https://doi.org/10.4049/jimmunol.175.1.591.

136. Blicher T, Kastrup JS, Buus S, Gajhede M. High-resolution structure of HLA-A*1101 in complex with SARS nucleocapsid peptide. Acta Crystallogr D Biol Crystallogr. 2005;61:1031–40 https://scripts.iucr.org/cgi-bin/paper?dz5040.

137. Rivino L, Tan AT, Chia A, Kumaran EAP, Grotenbreg GM, MacAry PA, et al. Defining CD8+ T cell determinants during human viral infection in populations of Asian ethnicity. J Immunol. 2013;191:4010–9. https://doi.org/10.4049/jimmunol.1301507.

138. Cheung YK, Cheng SCS, Sin FWY, Chan KT, Xie Y. Investigation of immunogenic T-cell epitopes in SARS virus nucleocapsid protein and their role in the prevention and treatment of SARS infection. Hong Kong Med J. 2008;14(Suppl 4):27–30 https://www.ncbi.nlm.nih.gov/pubmed/18708671.

139. Yang J, James E, Roti M, Huston L, Gebe JA, Kwok WW. Searching immunodominant epitopes prior to epidemic: HLA class II-restricted SARS-CoV spike protein epitopes in unexposed individuals. Int Immunol. 2009;21: 63–71. https://doi.org/10.1093/intimm/dxn124.

140. Yang L, Peng H, Zhu Z, Li G, Huang Z, Zhao Z, et al. Persistent memory CD4+ and CD8+ T-cell responses in recovered severe acute respiratory syndrome (SARS) patients to SARS coronavirus M antigen. J Gen Virol. 2007; 88(Pt 10):2740–8. https://doi.org/10.1099/vir.0.82839-0.

141. Poran A, Harjanto D, Malloy M, Rooney MS, Srinivasan L, Gaynor RB. Sequence-based prediction of vaccine targets for inducing T cell responses to SARS-CoV-2 utilizing the bioinformatics predictor RECON. doi:https://doi.org/10.1101/2020.04.06.027805.

142. Peng H, Yang L-T, Wang L-Y, Li J, Huang J, Lu Z-Q, et al. Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients. Virology. 2006;351:466–75. https://doi.org/10.1016/j.virol.2006.03.036.

143. Zhou M, Xu D, Li X, Li H, Shan M, Tang J, et al. Screening and identification of severe acute respiratory syndrome-associated coronavirus-specific CTL epitopes. J Immunol. 2006;177:2138–45. https://doi.org/10.4049/jimmunol.177.4.2138.

144. Kohyama S, Ohno S, Suda T, Taneichi M, Yokoyama S, Mori M, et al. Efficient induction of cytotoxic T lymphocytes specific for severe acute respiratory syndrome (SARS)-associated coronavirus by immunization with surface-linked liposomal peptides derived from a non-structural polyprotein 1a. Antiviral Res. 2009;84:168–77. https://doi.org/10.1016/j.antiviral.2009.09.004.

145. Libraty DH, O'Neil KM, Baker LM, Acosta LP, Olveda RM. Human CD4(+) memory T-lymphocyte responses to SARS coronavirus infection. Virology. 2007;368:317–21. https://doi.org/10.1016/j.virol.2007.07.015.

146. Calis JJA, Maybeno M, Greenbaum JA, Weiskopf D, De Silva AD, Sette A, et al. Properties of MHC class I presented peptides that enhance immunogenicity. Plos Comput Biol. 2013;9:e1003266. https://doi.org/10.1371/journal.pcbi.1003266.

147. Smith CC, Chai S, Washington AR, Lee SJ, Landoni E, Field K, et al. Machine-learning prediction of tumor antigen immunogenicity in the selection of therapeutic epitopes. Cancer Immunol Res. 2019;7:1591–604. https://doi.org/10.1158/2326-6066.CIR-19-0155.

148. Abelin JG, Keskin DB, Sarkizova S, Hartigan CR, Zhang W, Sidney J, et al. Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. Immunity. 2017;46:315–26. https://doi.org/10.1016/j.immuni.2017.02.007.

149. Davidson AD, Williamson MK, Lewis S, Shoemark D, Carroll MW, Heesom KJ, et al. Characterisation of the transcriptome and proteome of SARS-CoV-2 reveals a cell passage induced in-frame deletion of the furin-like cleavage site from the spike glycoprotein. Genome Med. 2020;12:68. https://doi.org/10.1186/s13073-020-00763-0.

150. Kim D, Lee J-Y, Yang J-S, Kim JW, Narry Kim V, Chang H. The architecture of SARS-CoV-2 transcriptome. doi:https://doi.org/10.1101/2020.03.12.988865.

151. Yates AD, Achuthan P, Akanni W, Allen J, Allen J, Alvarez-Jarreta J, et al. Ensembl 2020. Nucleic Acids Res. 2020;48:D682–8. https://doi.org/10.1093/nar/gkz966.

152. Grant OC, Montgomery D, Ito K, Woods RJ. 3D Models of glycosylated SARS-CoV-2 spike protein suggest challenges and opportunities for vaccine development. bioRxiv. 2020:2020.04.07.030445. doi:https://doi.org/10.1101/2020.04.07.030445.

153. Walls AC, Park YJ, Tortorici MA, Wall A. Seattle Structural Genomics Center for Infectious Disease (SSGCID), McGuire AT, et al. SARS-CoV-2 spike ectodomain structure (open state); 2020. https://doi.org/10.2210/pdb6vyb/pdb.

154. Watanabe Y, Allen JD, Wrapp D, McLellan JS, Crispin M. Site-specific analysis of the SARS-CoV-2 glycan shield. bioRxiv. 2020;:2020.03.26.010322. doi: https://doi.org/10.1101/2020.03.26.010322.

155. Xu Y, Zhu J, Liu Y, Lou Z, Yuan F, Liu Y, et al. Characterization of the heptad repeat regions, HR1 and HR2, and design of a fusion core structure model of the spike protein from severe acute respiratory syndrome (SARS) coronavirus. Biochemistry. 2004;43:14064–71. https://doi.org/10.1021/bi049101q.

156. Lai S-C, Chong PC-S, Yeh C-T, Liu LS-J, Jan J-T, Chi H-Y, et al. Characterization of neutralizing monoclonal antibodies recognizing a 15-residues epitope on the spike protein HR2 region of severe acute respiratory syndrome coronavirus (SARS-CoV). J Biomed Sci. 2005;12:711–27. https://doi.org/10.1007/s11373-005-9004-3.

157. He Y, Zhu Q, Liu S, Zhou Y, Yang B, Li J, et al. Identification of a critical neutralization determinant of severe acute respiratory syndrome (SARS)-associated coronavirus: importance for designing SARS vaccines. Virology. 2005;334:74–82. https://doi.org/10.1016/j.virol.2005.01.034.

158. He Y, Zhou Y, Liu S, Kou Z, Li W, Farzan M, et al. Receptor-binding domain of SARS-CoV spike protein induces highly potent neutralizing antibodies: implication for developing subunit vaccine. Biochem Biophys Res Commun. 2004;324:773–81. https://doi.org/10.1016/j.bbrc.2004.09.106.

159. Hu H, Li L, Kao RY, Kou B, Wang Z, Zhang L, et al. Screening and identification of linear B-cell epitopes and entry-blocking peptide of severe acute respiratory syndrome (SARS)-associated coronavirus using synthetic overlapping peptide library. J Comb Chem. 2005;7:648–56. https://doi.org/10.1021/cc0500607.

160. Madu IG, Roth SL, Belouzard S, Whittaker GR. Characterization of a highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. J Virol. 2009;83:7411–21. https://doi.org/10.1128/JVI.00079-09.

161. Kreiter S, Vormehr M, van de Roemer N, Diken M, Löwer M, Diekmann J, et al. Mutant MHC class II epitopes drive therapeutic immune responses to cancer. Nature. 2015;520:692–6. https://doi.org/10.1038/nature14426.

162. Iiizumi S, Ohtake J, Murakami N, Kouro T, Kawahara M, Isoda F, et al. Identification of novel HLA class II-restricted neoantigens derived from driver mutations. Cancers . 2019;11. doi:https://doi.org/10.3390/cancers11020266.

163. Bekri S, Uduman M, Gruenstein D, Mei AH-C, Tung K, Rodney-Sandy R, et al. Neoantigen synthetic peptide vaccine for multiple myeloma elicits T cell immunity in a pre-clinical model. Blood. 2017;130(Supplement 1):1868. https://doi.org/10.1182/blood.V130.Suppl_1.1868.1868.

Smith *et al. Genome Medicine*     (2021) 13:101

Page 23 of 23

164. Ferretti AP, Kula T, Wang Y, Nguyen DMV, Weinheimer A, Dunlap GS, et al. Unbiased screens show CD8 T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that Largely Reside outside the Spike Protein. Immunity. 2020;53:1095–107.e3. https://doi.org/10.101 6/j.immuni.2020.10.006.

165. Klinger M, Pepin F, Wilkins J, Asbury T, Wittkop T, Zheng J, et al. Multiplex Identification of antigen-specific T cell receptors using a combination of immune assays and immune receptor sequencing. Plos One. 2015;10: e0141561. https://doi.org/10.1371/journal.pone.0141561.

166. Reiss S, Baxter AE, Cirelli KM, Dan JM, Morou A, Daigneault A, et al. Comparative analysis of activation induced marker (AIM) assays for sensitive identification of antigen-specific CD4 T cells. Plos One. 2017;12:e0186998. https://doi.org/10.1371/journal.pone.0186998.

167. Wang Z, Cheng G, Li G. TCR Ligand Discovery via T-Scan. Trends Immunol. 2019;40:1075–7. https://doi.org/10.1016/j.it.2019.10.003.

168. Amanat F, Krammer F. SARS-CoV-2 Vaccines: Status Report. Immunity. 2020. https://doi.org/10.1016/j.immuni.2020.03.007.

169. Wang L, Shi W, Chappell JD, Joyce MG, Zhang Y, Kanekiyo M, et al. Importance of neutralizing monoclonal antibodies targeting multiple antigenic sites on the Middle East respiratory syndrome coronavirus spike glycoprotein to avoid neutralization escape. J Virol. 2018;92. https://doi. org/10.1128/JVI.02002-17.

170. Du L, Tai W, Yang Y, Zhao G, Zhu Q, Sun S, et al. Introduction of neutralizing immunogenicity index to the rational design of MERS coronavirus subunit vaccines. Nat Commun. 2016;7:13473. https://doi.org/1 0.1038/ncomms13473.

171. Li Y, Wan Y, Liu P, Zhao J, Lu G, Qi J, et al. A humanized neutralizing antibody against MERS-CoV targeting the receptor-binding domain of the spike protein. Cell Res. 2015;25:1237–49. https://doi.org/10.1038/cr.2015.113.

172. Coleman CM, Liu YV, Mu H, Taylor JK, Massare M, Flyer DC, et al. Purified coronavirus spike protein nanoparticles induce coronavirus neutralizing antibodies in mice. Vaccine. 2014;32:3169–74. https://doi.org/10.1016/j.va ccine.2014.04.016.

173. Escriou N, Callendret B, Lorin V, Combredet C, Marianneau P, Février M, et al. Protection from SARS coronavirus conferred by live measles vaccine expressing the spike glycoprotein. Virology. 2014;452-453:32–41. https://doi. org/10.1016/j.virol.2014.01.002.

174. Ishii K, Hasegawa H, Nagata N, Ami Y, Fukushi S, Taguchi F, et al. Neutralizing antibody against severe acute respiratory syndrome (SARS)-coronavirus spike is highly effective for the protection of mice in the murine SARS model. Microbiol Immunol. 2009;53:75–82. https://doi.org/1 0.1111/j.1348-0421.2008.00097.x.

175. Kuate S, Cinatl J, Doerr HW, Uberla K. Exosomal vaccines containing the S protein of the SARS coronavirus induce high levels of neutralizing antibodies. Virology. 2007;362:26–37. https://doi.org/10.1016/j.virol.2006.12.011.

176. Woo PCY, Lau SKP, Tsoi H-W, Chen Z-W, Wong BHL, Zhang L, et al. SARS coronavirus spike polypeptide DNA vaccine priming with recombinant spike polypeptide from Escherichia coli as booster induces high titer of neutralizing antibody against SARS coronavirus. Vaccine. 2005;23:4959–68. https://doi.org/10.1016/j.vaccine.2005.05.023.

177. Grifoni A, Sidney J, Zhang Y, Scheuermann RH, Peters B, Sette A. A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe. 2020;27: 671–80.e2. https://doi.org/10.1016/j.chom.2020.03.002.

178. Forsström B, Axnäs BB, Rockberg J, Danielsson H, Bohlin A, Uhlen M. Dissecting antibodies with regards to linear and conformational epitopes. Plos One. 2015;10:e0121673. https://doi.org/10.1371/journal.pone.0121673.

179. Van Regenmortel MHV. What is a B-cell epitope? In: Epitope Mapping Protocols. Springer; 2009. p. 3–20. https://link.springer.com/protocol/10.1 007/978-1-59745-450-6_1. Accessed 21 May 2020.

180. Ito HO, Nakashima T, So T, Hirata M, Inoue M. Immunodominance of conformation-dependent B-cell epitopes of protein antigens. Biochem Biophys Res Commun. 2003;308:770–6. https://doi.org/10.1016/s0006-291 x(03)01466-9.

181. Poran A, Harjanto D, Malloy M, Arieta CM, Rothenberg DA, Lenkala D, et al. Sequence-based prediction of SARS-CoV-2 vaccine targets using a mass spectrometry-based bioinformatics predictor identifies immunogenic T cell epitopes. Genome Med. 2020;12:70. https://doi.org/10.1186/s13073-020-00767-w.

182. Liu G, Carter B, Bricken T, Jain S, Viard M, Carrington M, et al. Computationally optimized SARS-CoV-2 MHC class I and II vaccine

183. Yarmarkovich M, Warrington JM, Farrel A, Maris JM. Identification of SARS-CoV-2 vaccine epitopes predicted to induce long-term population-scale immunity. Cell Rep Med. 2020;1:100036. https://doi.org/10.1016/j. xcrm.2020.100036.

184. O'Donnell TJ, Rubinsteyn A, Laserson U. MHCflurry 2.0: improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. Cell Syst. 2020;11:418–9. https://doi.org/10.1016/j.cels.2020.09.001.

185. Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR, et al. The immune epitope database (IEDB) 3.0. Nucleic Acids Res. 2015; 43(Database issue):D405–12. https://doi.org/10.1093/nar/gku938.

186. Gao A, Chen Z, Segal FP, Carrington M, Streeck H, Chakraborty AK, et al. Predicting the Immunogenicity of T cell epitopes: From HIV to SARS-CoV-2. doi:https://doi.org/10.1101/2020.05.14.095885.

187. Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature. 2017;551:517–20. https://doi.org/10.1038/ nature24473.

188. Xu Z, Shi L, Wang Y, Zhang J, Huang L, Zhang C, et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir Med. 2020;8:420–2. https://doi.org/10.1016/S2213-2600(20)30076-X.

189. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395:497–506. https://doi.org/10.1016/S0140-6736(20)30183-5.

190. Faure E, Poissy J, Goffard A, Fournier C, Kipnis E, Titecat M, et al. Distinct immune response in two MERS-CoV-infected patients: can we go from bench to bedside? Plos One. 2014;9:e88716. https://doi.org/10.1371/journal. pone.0088716.

191. Wu D, Yang XO. TH17 responses in cytokine storm of COVID-19: an emerging target of JAK2 inhibitor Fedratinib. J Microbiol Immunol Infect. 2020;53:368–70. https://doi.org/10.1016/j.jmii.2020.03.005.

192. Shi G, Vistica BP, Nugent LF, Tan C, Wawrousek EF, Klinman DM, et al. Differential involvement of Th1 and Th17 in pathogenic autoimmune processes triggered by different TLR ligands. J Immunol. 2013;191:415–23. https://doi.org/10.4049/jimmunol.1201732.

193. Jyotisha, Singh S, Qureshi IA. Multi-epitope vaccine against SARS-CoV-2 applying immunoinformatics and molecular dynamics simulation approaches. J Biomol Struct Dyn. 2020;1–17. doi:https://doi.org/10.1080/073 91102.2020.1844060.

194. Behmard E, Soleymani B, Najafi A, Barzegari E. Immunoinformatic design of a COVID-19 subunit vaccine using entire structural immunogenic epitopes of SARS-CoV-2. Sci Rep. 2020;10:20864. https://doi.org/10.1038/s41598-020-77547-4.

195. Kwarteng A, Asiedu E, Sakyi SA, Asiedu SO. Targeting the SARS-CoV2 nucleocapsid protein for potential therapeutics using immuno-informatics and structure-based drug discovery techniques. Biomed Pharmacother. 2020;132:110914. https://doi.org/10.1016/j.biopha.2020.11 0914.

196. Le Bert N, Tan AT, Kunasegaran K, Tham CYL. Different pattern of pre-existing SARS-COV-2 specific T cell immunity in SARS-recovered and uninfected individuals. bioRxiv. 2020. https://www.biorxiv.org/content/10.11 01/2020.05.26.115832v1.abstract. Accessed 8 May 2021.

197. Quadeer AA, Ahmed SF, McKay MR. Epitopes targeted by T cells in convalescent COVID-19 patients. bioRxiv. 2020;:2020.08.26.267724. doi: https://doi.org/10.1101/2020.08.26.267724.

198. Ou L, Kong W-P, Chuang G-Y, Ghosh M, Gulla K, O'Dell S, et al. Preclinical Development of a Fusion Peptide Conjugate as an HIV Vaccine Immunogen. Sci Rep. 2020;10:3032. https://doi.org/10.1038/s41598-020-59711-y.

199. Smith CC, Rubinsteyn A. Landscape-and-Selection-of-Vaccine-Epitopes-in-SARS-CoV-2. Github. 2021. https://github.com/Benjamin-Vincent-Lab/La ndscape-and-Selection-of-Vaccine-Epitopes-in-SARS-CoV-2. Accessed 8 May 2021.

200. Smith CC. Landscape and selection of vaccine epitopes in SARS-CoV-2. Mendeley Data, V6; 2021. https://doi.org/10.17632/c6pdfrwxgj.6.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.