Title:

T cell epitope mapping reveals immunodominance of evolutionarily conserved regions within SARS-CoV-2 proteome.

Authors:

Cansu Cimen Bozkus[1,2,3†*], Matthew Brown[1,2,3,4†], Leandra Velazquez[1,2], Marcus Thomas[5], Eric A. Wilson[1,5,6], Timothy O'Donnell[1,2], Denis Ruchnewitz[7], Douglas Geertz[1,2], Yonina Bykov[4,8], Julia Kodysh[1,2], Kasopefoluwa Y. Oguntuyo[9], Vladimir Roudko[5], David Hoyos[10], Komal D. Srivastava[8,11], Giulio Kleiner[8,11], Hala Alshammary[8,11], Neha Karekar[1,2], Christopher McClain[1,2], Ramya Gopal[1,2], Kai Nie[12], Diane Del Valle[4], Daniela Delbeau-Zagelbaum[1], Denise Rodriguez[1], Jessica Setal[1], The Mount Sinai COVID-19 Biobank Team, Emily Carroll[13], Margrit Wiesendanger[13], Percio S. Gulko[13], Alexander Charney[14], Miriam Merad[5,6], Seunghee Kim-Schulze[5,12], Benhur Lee[8], Ania Wajnberg[9], Viviana Simon[8,11,15,16,17], Benjamin D Greenbaum[10,18], Diego Chowell[1,5,6], Nicolas Vabret[5,6], Marta Luksza[5,14], Nina Bhardwaj[1,2,3*]

Affiliations:

[1]Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[2]The Department of Medicine, The Division of Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[3]Parker Institute of Cancer Immunotherapy, San Francisco, CA, USA.

[4]Graduate School of Biomedical Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[5]Department of Oncological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[6]The Department of Immunology and Immunotherapy, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[7]Institute for Biological Physics, University of Cologne, 50937 Cologne, Germany.

[8]Department of Microbiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[9]The Department of Medicine, The Division of General Internal Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[10]Computational Oncology, Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA.

[11]Center for Vaccine Research and Pandemic Preparedness (C-VaRPP), Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[12]Human Immune Monitoring Center, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[13]The Department of Medicine, The Division of Rheumatology, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[14]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[15]Department of Pathology, Molecular and Cell Based Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[16]The Department of Medicine, The Division of Infectious Diseases, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[17]The Global Health and Emerging Pathogens Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA.

[18]Physiology, Biophysics & Systems Biology, Weill Cornell Medicine, Weill Cornell Medical College, New York, NY, USA

†These authors contributed equally to this work.
*Corresponding authors. Email: cansu.cimenbozkus@mssm.edu, nina.bhardwaj@mssm.edu

Abstract:

As SARS-CoV-2 variants continue to emerge capable of evading neutralizing antibodies, it has become increasingly important to fully understand the breadth and functional profile of T cell responses to determine their impact on the immune surveillance of variant strains. Here, sampling healthy individuals, we profiled the kinetics and polyfunctionality of T cell immunity elicited by mRNA vaccination. Modeling of anti-spike T cell responses against ancestral and variant strains of SARS-CoV-2 suggested that epitope immunodominance and cross-reactivity are major predictive determinants of T cell immunity. To identify immunodominant epitopes across the viral proteome, we generated a comprehensive map of $CD4^+$ and $CD8^+$ T cell epitopes within non-spike proteins that induced polyfunctional T cell responses in convalescent patients. We found that immunodominant epitopes mainly resided within regions that were minimally disrupted by mutations in emerging variants. Conservation analysis across historical human coronaviruses combined with *in silico* alanine scanning mutagenesis of non-spike proteins underscored the functional importance of mutationally-constrained immunodominant regions. Collectively, these findings identify immunodominant T cell epitopes across the mutationally-constrained SARS-CoV-2 proteome, potentially providing immune surveillance against emerging variants, and inform the design of next-generation vaccines targeting antigens throughout SARS-CoV-2 proteome for broader and more durable protection.

One Sentence Summary:

Polyfunctional CD8+ and CD4+ T cells directed against SARS-CoV-2 target mutationally constrained regions of the viral proteome.

Main Text:

INTRODUCTION

Since its initial emergence in December 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) genome has consistently undergone mutations, resulting in new variants. Some of these variants, Alpha, Beta, Gamma, Delta and Omicron, spread rapidly, causing outbreaks, and are designated as "variants of concern (VOC)". VOC are marked by their enhanced transmissibility, effectively outcompeting their predecessors due to increased viral fitness and escape from immune recognition (*1*). Although the COVID-19 vaccines are continuously adapted to target prevalent VOC, by the time they are commonly accessible, the next variant will likely have emerged, compromising the vaccine's efficacy. This race between the new variants and effective vaccine development warrants the design of next-generation vaccines that can provide unfettered protection against future SARS-CoV-2 variants.

Coordinated antigen-specific B and T cell responses, induced by vaccination and/or infection, are imperative to clearing RNA viruses (*2*, *3*). Accordingly, the role of neutralizing antibodies in preventing SARS-CoV-2 infection (*4*) and improving disease outcomes has been well documented (*5*). However, antibody titers wane over time (*6*) and VOC frequently evade neutralization by antibodies (*7–9*). Despite reports of certain mutations evading T cell recognition (*10*, *11*), there is accumulating evidence that demonstrates preservation of overall T cell responses against VOC by targeting conserved prevalent epitopes (*12*) or cross-recognizing mutated epitopes (*13*). Notably, T cell responses are critical for controlling SARS-CoV-2 infection and SARS-CoV-2-specific CD4+ and CD8+ T cells correlate with durable immune protection and reduced disease severity (*14–16*). Corroborating the durability of T cell immunity, COVID-19 vaccines have been shown to induce stem-like memory T cells persisting more than 6 months after vaccination (*17*). Additionally, in preclinical studies using transgenic mice infected with SARS-CoV-2 (*18*) and rhesus macaques administered intranasal SARS-CoV-2 vaccines, CD8+ T cells alone were shown to be adequate for viral clearance even in the absence of humoral immunity (*19*). Furthermore, in patients receiving B cell-depleting therapies, vaccines elicited sustained T cells responses against SARS-CoV-2, providing protection from severe disease (*20*, *21*). Therefore, T cells may be key to mediating long-lasting, protective immunity against SARS-CoV-2 amid emerging variants that can evade antibody responses.

Detailed interrogation of T cell specificities from patient cohorts exposed to SARS-CoV-2 has demonstrated the induction of T cell responses against a range of viral proteins, among which spike (S), nucleocapsid (N), membrane (M), ORF1, and ORF3 appear to be dominant targets (*22–25*). Additionally, studies of unexposed individuals identified preexisting memory T cells recognizing SARS-CoV-2 sequences (*22*, *26*, *27*), which are reported to be cross-reactive T cells derived from previous exposure to common cold viruses (HCoVs) (*28*). HCoVs share partial sequence homology to SARS-CoV-2, across both structural and non-structural proteins, (*29*) and circulate widely, having infected a significant proportion of the population (*30*, *31*). A potential protective role for HCoV-derived cross-reactive T cells has also been suggested in individuals subsequently exposed to SARS-CoV-2 (*32*). Thus, incorporating immunogenic T cell epitopes from conserved regions of the SARS-CoV-2 proteome into immunization strategies may offer durable and comprehensive protection, as reinforcement against emerging variants that escape humoral immunity (*33*).

113 Here, we first investigated the specificities of T cell responses towards immunogenic regions in
114 S that are subject to mutational events in SARS-CoV-2 VOC in cohorts of vaccinated
115 individuals. Our kinetics studies confirmed that T cells were induced early on after mRNA
116 vaccination with a high frequency of polyfunctional populations able to contribute to coordinated
117 adaptive immunity. To understand the determinants governing the preservation of T cell
118 responses and viral immune escape, we extensively mapped the functional responses to mutated
119 S epitopes revealing that polyclonal T cell populations were preserved and cross-reactive against
120 variant antigens. Additionally, because internal antigens also serve as potent targets of T cell
121 immunity and experience less antibody-driven mutational pressure compared to S, we conducted
122 one of the largest experimental efforts to-date for combined CD4+ and CD8+ T cell epitope
123 mapping of non-S proteins. Our comprehensive approach revealed highly immunogenic,
124 dominant T cell epitopes across a cohort of convalescent patients, eliciting conserved cellular
125 immune responses, which could serve as candidate epitopes for incorporation into vaccine
126 designs effective against emerging SARS-CoV-2 variants and potentially pan-coronavirus
127 vaccine strategies.

128 **RESULTS**

129 **Adaptive immune responses elicited by SARS-CoV-2 mRNA vaccination**

130 To study SARS-CoV-2 mRNA vaccine-induced adaptive immunity, we collected longitudinal
131 blood samples from 15 healthy donors, who received either mRNA-1273 (by Moderna) or
132 BNT162b2 (by Pfizer/BioNTech) vaccine series (table S1). Samples were collected before
133 vaccination (V0), 14 days after 1st dose (V1D14), 7 days after 2nd dose (V2D7), and 14 days
134 after 2nd dose (V2D14) (Fig. 1A). We assessed the kinetics of the neutralizing antibody
135 responses, a major correlate of protection from SARS-CoV-2 (*4*), using a pseudotyped vesicular
136 stomatitis virus (VSVΔG-Rluc) expressing the D614G SARS-CoV-2 spike (S) protein.
137 Neutralizing antibody titers were significantly increased after a single dose of mRNA vaccine
138 and continued to increase after the second dose exceeding the titer levels detected in most
139 unvaccinated convalescent control (CC) patients (Fig. 1B and fig. S1). Notably, 2 of the
140 individuals in the vaccination cohort were previously infected with SARS-CoV-2. A single dose
141 of mRNA vaccine was sufficient to robustly boost neutralizing antibody titers in these
142 individuals (Fig. 1B).

143 Next, we interrogated the kinetics of vaccine-induced T cell responses, another key correlate of
144 protection from SARS-CoV-2 (*14*). Peripheral blood mononuclear cells (PBMCs) collected from
145 vaccinated donors at V0, V1D14, V2D7, and V2D14 were stimulated in an *ex vivo* IFN-γ
146 ELISpot assay with pooled 15mer overlapping peptides (OLPs) spanning the N- and C-terminal
147 halves of Wuhan-1 S protein, Spike_N and Spike_C, respectively (table S2). mRNA vaccination
148 significantly induced anti-S T cell responses at all timepoints tested, demonstrating *in vivo*
149 priming. However, there was no significant change in the magnitude of T cell responses across
150 post-vaccine timepoints tested (Fig. 1C). Notably, the magnitude of anti-S T cell responses
151 elicited by mRNA-1273 was significantly greater than those elicited by BNT162b2 vaccine (Fig.
152 1C). This may be due to different antigen loads and dosing schedules of the two mRNA vaccines
153 (*34*). Similar differences between mRNA-1273 and BNT162b2-induced humoral responses were
154 previously reported (*34*).

155 To study the abundance and functionality of vaccine-induced T cells more robustly, we expanded
156 Wuhan-1 S-specific T cells in pre- and post-vaccination blood collected at V0 and V2D14,

respectively. T cells were stimulated with Spike_N and Spike_C OLP pools and antigen-specific T cells were expanded for 10 days before measuring effector cytokine production by intracellular flow cytometry (*35*). Vaccination induced robust anti-S T cell responses in all donors tested. Although less frequently, anti-S T cells were detected at V0 in some donors (Fig. 1D). This can be due to *in vitro* priming of naïve T cells (*35*) and/or expansion of cross-reactive T cells. mRNA vaccines could elicit both CD4+ and CD8+ anti-S T cells; however, reactive T cells were predominantly CD4+ in most of the donors (Fig. 1D). This was consistent with our observations in unvaccinated convalescent donors (CPC cohort, table S1), where infection-induced anti-S T cells were also predominantly CD4+ (Fig. 1E). As expected by their unique major histocompatibility complex (MHC) molecules (table S1) and T cell repertoires, the abundance and CD4:CD8 subset distributions of anti-S T cells varied across donors. Importantly, vaccines could induce polyfunctional CD4+ and CD8+ anti-S T cells secreting IFN-γ, TNF-α and IL-2 (Fig. 1F and G), further supporting the role of mRNA vaccination in establishing an effective T cell immunity against SARS-CoV-2. Induction of polyfunctional anti-S T cells was similar between mRNA-1273 and BNT162b2 recipients (fig. S2).

Cross-reactivity of vaccine-induced adaptive immunity against SARS-CoV-2 variants

Breakthrough infections caused by emerging variants of concern indicate an impaired ability of the vaccine-induced immunity to recognize SARS-CoV-2 variants. To assess the kinetics of Wuhan-1 S-directed antibody binding to variant strains, we utilized sera collected at V0, V1D14, V2D7, and V2D14 in a Luminex binding assay with multiple spike receptor binding domain (RBD) constructs. Antibody recognition of ancestral and Alpha (B.1.1.7) strain RBD was comparable across all timepoints tested, while there was a reduction in the recognition of Beta (B.1.351) and Gamma (P.1) RBD (Fig. 2A), consistent with other studies showing limited antigenic change with Alpha (*36*) and reduced viral neutralization with Beta and Gamma variants (*37*). The reduced recognition of Beta and Gamma RBD was even more pronounced in convalescent patients and was at least in part mediated by E484K mutation found in both strains with potential contributions by the K417N mutation found in Beta, Delta and Omicron strains (Fig. 2A).

Cross-reactive T cell responses are also important for controlling infections with SARS-CoV-2 variant strains, especially given the diminished neutralizing antibody control. To study cross-reactive T cell responses against S mutations identified in variants of concern, we designed 15mer OLPs spanning a maximum of 14 upstream and downstream amino acid (aa) sequences surrounding each mutated aa. Then we pooled the WT (Wuhan-1) and mutant OLPs corresponding to each variant tested, namely Alpha, Beta, and Gamma (table S2 and fig. S3). Using these pools, we stimulated PBMCs collected from vaccinated (V2D14) and convalescent (unvaccinated) individuals in an *ex vivo* IFN-γ ELISpot assay to measure the T cell recognition of WT and variant S. Both vaccine and infection-primed Wuhan-1 S-specific T cells recognized mutated S in Alpha, Beta and Gamma variants at a similar capacity *ex vivo* (Fig. 2B). These observations are in accordance with previous studies demonstrating preservation of overall T cell responses against early variants in *ex vivo* assays (*12*, *13*). However, multiple studies have reported that certain S mutations may lead to escape from T cell recognition (*10*, *11*). It is possible that *ex vivo* analysis of T cell cross-reactivity may not capture nuanced differences in the antigenicity of S mutations since total anti-S T cells comprise a small proportion among PBMCs and the mutated S epitopes tested here are only a subset of many immunogenic S epitopes engendering a wide breadth of T cell responses (*38*). Therefore, we aimed to study the

202 antigenicity and cross-recognition of S mutations in clonally enriched T cell populations specific
203 to these epitope regions. We expanded T cells from vaccinated donors collected at V2D14
204 stimulated with variant peptide pools and the corresponding WT pools, as described for *ex vivo*
205 analysis. To measure antigen-specific effector responses, expanded T cells were re-stimulated
206 with either the WT or mutant peptides with which they were initially stimulated. In addition,
207 expanded WT T cells were re-stimulated with the corresponding mutant epitopes to directly
208 measure the cross-recognition of mutant S epitopes by Wuhan-1 S-specific T cells. In accordance
209 with total S-specific T cell immunity measured in fig. 1D, WT and mutated S-specific reactive T
210 cells were predominantly CD4+ (Fig. 2C). Both WT and mutated S in Alpha, Beta, and Gamma
211 variants induced overall T cell reactivity similarly (Fig. 2C). However, the direct measurement of
212 cross-recognition revealed a significantly diminished reactivity of WT T cells against variant S
213 (Fig. 2C).

214 To evaluate the underlying parameters of T cell immunity, we modeled anti-S T cell responses
215 against ancestral and variant strains of SARS-CoV-2 measured in fig. 2C. Our basic model,
216 detailed in the methods section, considers T cell recognition as a two-step process: 1) peptide
217 binding and presentation by MHC (pMHC) and 2) binding of pMHCs by specific T cell
218 receptors (TCRs), and includes the effects of initial stimulation, re-stimulation, and peptide
219 cross-reactivity, as well as patient specific amplitude of immune response. Our pMHC
220 calculations utilize prediction algorithms, which perform better for MHC-I than MHC-II binding
221 (*39*). Therefore, we focused on CD8+ T cell responses. We computed a recognition score for
222 each 9mer within the peptide pools over the MHC-I set for each donor. To evaluate the
223 contribution of 9mers to overall pool level response, we applied different 9mer aggregating
224 functions, including "*max*", "*sum*", and biologically unrealistic "*min*" as a control. The *max*
225 function assumes that the most immunogenic 9mer in a pool is largely responsible for the
226 observed T cell response, while the *sum* function assumes all 9mers contribute to the pool-level
227 response in proportion to their relative immunogenicity. To compare different aggregating
228 functions, we derived the Bayesian information criterion (BIC) values for each, where a lower
229 value is preferred. The *max* function showed the best performance (Fig. 2D and fig. S4),
230 suggesting that anti-S T cell responses are primarily mounted by immunodominant epitopes.
231 Expectedly, the control *min* function, which assumes the least immunogenic epitope would drive
232 the observed response, performed poorly. Further supporting the immunodominance model over
233 the additive model, the *max* and *sum* functions performed similarly since the recognition scores
234 of 9mers in a peptide pool were not uniform but typically dominated by a small number of 9mers
235 with negligible scores for the remaining 9mers (Fig. 2D and fig. S5).

236 For the immunodominance model, we also compared extended and partial recognition models.
237 Introducing a term accounting for the sequence similarity of our test epitopes and immunogenic
238 epitopes in the immune epitope database (IEDB) slightly improved model performance in
239 accordance with previous reports attributing this approach to increased TCR response (*40*, *41*).
240 Removing patient-specific immune response amplitudes by setting them all to the same
241 optimized value displayed significantly reduced performance (Fig. 2D), highlighting the
242 heterogeneity of anti-S T cell responses observed across different donors. Our basic model
243 accounts for cross-reactivity of peptides, computing their sequence-based distance and resulting
244 impact on binding strength to the same TCR (*42*). Because all patients in our cohort were
245 vaccinated, we included the vaccine as a zeroth stimulation event, introducing additional cross-
246 reactivity terms with the complete Wuhan-1 S 9mers. However, we did not observe any
247 improvement in the model's performance, likely because the experimental data reflects only a

248 small subset of anti-S T cells and *in vitro* expansion impacts clonal dynamics. To evaluate the
249 role of cross-reactivity in variant S recognition by WT T cells more directly, we excluded the
250 effect of initial stimulation and consequently observed a decreased performance (Fig. 2D).

251 Together these results suggest that immunodominant epitopes are key inducers of anti-S T cell
252 immunity and that the overall anti-S T cell responses are preserved against mutant variants. As
253 the direct measurement of cross-recognition suggested a diminished reactivity of WT CD4+ T
254 cells against variant S, we investigated whether any specific S mutation may lead to escape from
255 T cell recognition. Stimulation of vaccinated donor (V2D14) T cells with peptides spanning
256 individual mutations found in variant S and the corresponding WT demonstrated a significant
257 decrease across the population in the recognition of mutations P26S and R246I by CD4+ T cells
258 (Fig. 2E). Although the impact of individual variant spike mutations on the T cell immunity is
259 likely compensated by the wide breadth of T cell responses elicited in vaccinated individuals, the
260 reduced cross-recognition of variant S and immune escape at certain mutant S epitopes prompt
261 the inclusion of non-spike epitopes in future vaccine designs to further boost T cell immunity
262 against SARS-CoV-2.

263 **T cell epitope mapping of non-spike SARS-CoV-2 proteins in convalescent patients**

264 To evaluate the immunogenicity of non-spike epitopes across SARS-CoV-2 proteome, we
265 synthesized overlapping 15mer peptides covering the entire nucleocapsid (N) and selected
266 regions in other proteins (Fig. 3A and table S2). Selected regions were determined by prioritizing
267 those enriched in interactions between epitopes predicted to be strong binders (*43*) and MHC
268 alleles frequent across different races (*44*) (fig. S6). We measured T cell responses in
269 unvaccinated convalescent donors (ATLAS cohort, table S1) utilizing *ex vivo* IFN-γ ELISpot and
270 T cell expansion assays (Fig. 3B). Both S and non-S SARS-CoV-2 proteins elicited robust T cell
271 responses *ex vivo* in convalescent patients (Fig. 3C) compared to healthy donor controls sampled
272 prior to COVID-19 pandemic (fig. S7). The distribution of memory T cell responses against each
273 protein varied across patients with S, N, membrane (M) and ORF7-directed responses being the
274 most potent (Fig. 3D). Expansion assays confirmed the immunogenicity of non-S proteins and
275 demonstrated that like anti-S T cells, non-S-specific T cells were also predominantly CD4+ (Fig.
276 3, E and F).

277 Since our modeling of anti-S T cell responses suggested that T cell immunity was primarily
278 elicited by immunodominant epitopes, we evaluated the presence of immunodominant regions
279 within non-S proteins. We stimulated convalescent donor T cells with individual 15mer peptides
280 constituting the peptide pools used in fig. 3. Epitope mapping revealed immunogenic regions
281 across non-S proteins that elicited polyfunctional CD4+ and CD8+ T cell responses (Fig. 4 and
282 table S3). All mapped non-S proteins contained multiple immunogenic epitopes. A peptide was
283 considered immunogenic, a "hit", if in at least one patient, the percent of reactive cells was
284 greater than the paired DMSO percent plus 3 times the standard deviation of all DMSO values
285 (>DMSO+3SD) across the population. We investigated the novelty of these non-S immunogenic
286 "hit" peptides by predicting the minimal epitopes for each "hit" binding to the responsive
287 patients' MHC alleles. A peptide was considered novel if none of its predicted binders were
288 deposited in IEDB. Accordingly, we found that 42% of CD8+ and 23% of CD4+ T cell peptides
289 we identified were novel, and the others were previously reported (fig. S8), providing additional
290 confidence in the validity of our epitope discovery pipeline. Notably, certain immunogenic
291 regions were commonly recognized by T cells across different convalescent patients (Fig. 4 and
292 table S3). Together, these data demonstrate the wide breadth and robustness of T cell responses

293 against non-S SARS-CoV-2 proteins and suggest that certain regions are immunodominant,
294 significantly inducing T cell immunity across the population.

295 **Conservation of non-spike immunodominant T cell epitopes**

296 A high rate of recurrent mutations is observed across all regions of the SARS-CoV-2 genome
297 (*45*), which can lead to escape from T cell immunosurveillance. Therefore, we examined the
298 mutational diversity within non-S proteins, especially in immunodominant regions. First, we
299 investigated the conservation of immunogenic ("hit") sequences, listed in fig. 4, among SARS-
300 CoV-2 variants by measuring the percentage of sequences that are deposited to GISAID and
301 have an exact match to the "hit" peptides. We observed a high degree of conservation with a
302 median over 99% for "hit" peptides inducing both CD4+ and CD8+ T cell responses (Fig. 5A).
303 To more closely investigate the mutational diversity across the non-S peptidome, we computed
304 the entropy of the observed amino acid frequencies on each codon position for each non-S
305 protein utilizing sequencing and regional epidemiological count data obtained from GISAID
306 EpiCoV database (*46*) and WHO Coronavirus dashboard (*47*), respectively, through different
307 timepoints (*48*). We then compared the entropies of codons encoding experimentally validated
308 immunogenic peptides ("hit" or immunodominant, table S3) vs those encoding the rest of the
309 non-S proteome. We found that codons encoding immunogenic peptides, especially those that
310 are dominant across the population, exhibit significantly lower entropy (Fig. 5B). To validate
311 this, we conducted a parallel analysis utilizing entropy data from Nextstrain. Consistently, our
312 findings confirmed the lower entropy values in immunogenic regions (fig. S9), thus further
313 demonstrating that immunogenic non-S sequences are derived from mutationally constrained
314 regions.

315 Next, we focused on nucleocapsid as it is the only non-S protein we epitope mapped across the
316 entire protein sequence, and its immunogenicity data lacks any potential bias introduced by our
317 initial peptide selection based on pMHC binding predictions. We aligned the normalized entropy
318 values per amino acid residue with heatmaps displaying the extent of sharing of immunogenic
319 residues across our cohort, alongside a map demonstrating regions enriched for immunogenic
320 residues even despite the scarcity of predicted strong binders. This enrichment signified the
321 production of immunogenic epitopes at a rate higher than would be suggested by the number of
322 predicted binders in regions of lower entropy that also contained shared immunogenic residues
323 (Fig. 5C). Importantly, these immunodominant, lower entropy regions mostly resided in highly
324 structured domains of nucleocapsid protein, namely the RNA binding domain (RBD) and
325 dimerization domain (DD) (Fig. 5C). This observation was not skewed by certain MHC alleles in
326 our cohort preferentially binding to highly structured domains, as previously reported in other
327 viral infections (*49*), since approximately 95% of all alleles were predicted to bind to epitopes
328 from RBD and DD more frequently than other, less structured domains (fig. S10). Finally, the
329 mapping of the immunogenic nucleocapsid epitopes reported in IEDB also showed an
330 enrichment within RBD and DD compared to less structured regions (fig. S11).

331 Together these observations led to the hypothesis that immunogenic epitopes are enriched in the
332 less diverse, functionally important regions of the SARS-CoV-2 proteome, thus likely exerting
333 evolutionary constraints limiting mutational rates to avoid compromised viral fitness.
334 Accordingly, we examined epitope conservation throughout coronavirus evolution by comparing
335 SARS-CoV-2 nucleocapsid amino acid sequence to other historical human coronaviruses,
336 namely HCoV-229E, HCoV-NL63, HCoV-HKU1, and HCoV-OC43. Corroborating previous
337 observations, immunogenic epitopes were more commonly found in fully conserved regions

338 compared to non-immunogenic residues (Fig. 5D and fig. S12-13). To further investigate if the
339 observed resistance to mutational pressure in the regions containing immunogenic epitopes is
340 related to maintaining protein function, we conducted *in silico* analyses focusing on RBD and
341 DD. First, we visualized the immunogenicity of each residue of RBD on its crystal structure and
342 found that immunogenic regions mapped onto the internal beta sheet, particularly the β-hairpin
343 (Fig. 5E), which is important for RNA binding (*50*, *51*). Then we tested whether the
344 immunogenic residues were important for RBD stability by mutating each residue to alanine and
345 comparing the energy of the modified structure to the wild type. The average change in protein
346 stability calculated for the top ten most immunogenic peptides from each patient was greater than
347 the overall average for residues across nucleocapsid RBD (Fig. 5F), indicating that most
348 immunogenic peptides contain structurally important residues. Similarly, for DD, immunogenic
349 regions mapped onto beta sheets (Fig. 5G), a critical motif in nucleocapsid oligomerization (*52*).
350 Furthermore, alanine scanning revealed that mutations in immunogenic peptides from all
351 patients, except Pt 003 which has missing data points, led to the destabilization of the protein-
352 protein interface (Fig. 5H), impacting nucleocapsid function.

353 Taken together, the mutational diversity and biophysical analyses indicate that epitopes inducing
354 T cell immunity reside in functionally and structurally important positions, potentially explaining
355 the high conservation and lack of immune editing in immunogenic regions.

356 **DISCUSSION**

357 Through the induction of adaptive immunity, vaccinations have markedly reduced SARS-CoV-2
358 infection rates, severe disease, and mortality. However, as antibodies wane over time and SARS-
359 CoV-2 evolves to evade host immunity, especially neutralizing antibodies, it is necessary to
360 develop new strategies that provide durable protection. Vaccine-induced S-specific T cell
361 responses have been shown to be long-lasting (*17*), overall preserved against variants (*12*, *13*)
362 and provide protection against COVID-19 even in the absence of antibody responses (*16*, *19–*
363 *21*). Unlike antibodies, T cell targets are not limited to extracellular antigens, making non-S viral
364 proteins a valuable source of potential T cell epitopes. Accumulating evidence demonstrates the
365 immunogenicity of non-S proteins, as well as their role in mediating viral clearance and disease
366 control (*22–25*, *38*, *53*, *54*). Thus, incorporating non-S proteins as targets in vaccine strategies
367 could enhance durable protection against emerging variants. Accordingly, several preclinical
368 studies have shown that inclusion of non-S antigens in immunization strategies provides robust
369 protection against VOC (*55–58*). In a hamster model of SARS-CoV-2 VOC challenge, a dual S
370 and N encoding mRNA vaccine, mRNA-S+N, had more robust control of Delta and Omicron
371 variants in the lungs and upper respiratory tract than mRNA-S. Although less robust than the
372 dual vaccine, mRNA-N alone was immunogenic and induced infection control (*57*). Moreover, a
373 recently developed mRNA vaccine, BNT162b4, encoding segments of N, M and ORF1ab
374 proteins, elicited T cell responses and provided protection against severe disease in murine
375 models when administered alone or in combination with S encoding BNT162b2. Protection
376 against VOC was enhanced with dual vaccination compared to vaccination with S encoding
377 vaccine alone (*58*). BNT162b4 is currently being tested in a clinical trial in combination with the
378 BNT162b2 Bivalent vaccine (NCT05541861).

379 As these new vaccines are designed, selection of epitope targets plays a critical role in achieving
380 maximum efficacy and protection against VOC across a diverse human population. Strategies for
381 selecting non-S targets may include utilizing entire proteins. Although this approach should

382 unbiasedly generate a broad range of epitopes, the high number of targets may cause immune
383 competition, potentially reducing the immunogenicity of key epitopes (*59*, *60*). Alternatively,
384 non-S targets may be selected by utilizing computational algorithms predicting immunogenic
385 epitopes. However, a high number of predicted epitopes does not elicit T cell responses (*61*). A
386 thorough understanding of SARS-CoV-2-specific T cell responses, informing their breadth,
387 immunodominance and conservation is needed to enable the design of broad-spectrum vaccines
388 effective against emerging variants.

389 Here we studied SARS-CoV-2-specific T cell responses from vaccine recipients and
390 convalescent patients via rigorous functional interrogation combined with *in silico* analyses. Our
391 work revealed the (i) kinetics of vaccination-induced polyfunctional CD4+ and CD8+ T cell
392 responses that remained intact against SARS-CoV-2 variants; (ii) immunodominant CD4+ and
393 CD8+ T cell epitopes across SARS-CoV-2 proteome; and that (iii) the immunodominant epitope
394 regions exist in protein domains of lower mutational diversity due at least in part to their
395 functional importance. Together our findings identify critical non-S epitopes that are conserved
396 and highly immunogenic across a diverse human cohort that can be utilized in future vaccine
397 designs to elicit a focused T cell response effective against emerging VOC.

398 Our studies evaluating mRNA vaccine-induced adaptive immunity response kinetics showed that
399 in SARS-CoV-2 infection-naïve individuals, while the antibody titers were peaked after the
400 booster vaccine, prime vaccination was sufficient to induce T cell responses, which remained
401 similarly high after the booster. This suggests that T cells may be key mediators of early
402 response to vaccination. Accordingly, Oberhardt *et al.* showed that functional CD8+ T cells were
403 mobilized even after 1 week of prime vaccination with BNT162b2 (*62*). Guerrera *et al.* also
404 showed the induction of T stem cell memory after prime vaccination, correlating with durable T
405 cell immunity (*17*). In our cohort the mRNA-1273 mRNA vaccine induced higher magnitude of
406 T cell responses than the BNT162b2 vaccine. Similar differences between the two vaccines were
407 reported for humoral responses (*34*). This is likely due different antigen loads and dosing
408 schedules of the two vaccines. Overall, both vaccines induced polyfunctional T cells and
409 similarly recognized tested VOC in *ex vivo* assays.

410 Our findings that while neutralizing antibodies diminish, T cell recognition of VOC is preserved
411 in *ex vivo* assays agree with previous reports (*12*, *13*) and underline the critical role of T cell
412 immunity in controlling emerging variants. This may be due to the wide breadth of T cell
413 epitopes such that a reduced capacity of T cell recognition of certain mutated epitopes is
414 compensated by the overall response. It may also be due to T cells that can cross-react with the
415 mutated epitopes. Our findings support the relevance of both mechanisms. Our results with
416 clonally enriched T cell populations specific to a single mutation region identified P26S and
417 R246I as escape mutations, significantly evading recognition by CD4+ T cells. P26S was also
418 previously reported to escape from CD4+ T cell responses (*13*). However, clonally enriched T
419 cell populations specific to multiple mutation regions found in each VOC clade showed no
420 difference in the overall recognition of mutant or ancestral sequences, compensating for any
421 diminished response to a single epitope. In addition, modeling of T cell responses against
422 ancestral and variant strains suggested cross-reactivity as a major predictive parameter of T cell
423 immunity. This mathematical modeling approach we utilized only involved CD8+ T cell
424 responses as it relied on peptide-MHC binding predictions, which perform poorly for MHC-II
425 binding. However, it is also important to understand CD4+ T cell dynamics as SARS-CoV-2
426 primarily induces CD4+ T cell responses and the reduction in cross-recognition of S mutations in

427 clonally enriched T cell populations was observed more robustly in the CD4+ T cell subset.
428 Despite cross-reactivity being a key parameter of T cell response to VOC, including vaccination
429 as an additional cross-reactivity measure with the complete Wuhan-1 S 9mers did not improve
430 the performance of our model. This is likely because our dataset is not well suited to measure the
431 vaccine's effect since our experimental dataset is focused on a specific subset of anti-S T cell
432 responses and *in vitro* expansion impacts clonal dynamics of T cells, reducing TCR diversity and
433 allowing *in* vitro priming which may introduce a new set of TCRs.

434 Modeling of anti-S T cell immunity also identified peptide immunodominance as a key
435 parameter. Others have also reported immunodominance, with a limited number of epitopes
436 accounting for most of the total response against SARS-CoV-2 (*24*, *38*). Here we
437 comprehensively mapped CD4+ and CD8+ T cell responses and found that patterns of
438 immunodominance were similarly present in non-S proteins, a subset of epitopes robustly
439 inducing T cell responses across a diverse cohort of patients. SARS-CoV-2 exhibits high rates of
440 recurrent mutations across its genome (*45*), often as a mechanism to evade immune responses.
441 This is observed particularly in VOC escaping from neutralizing antibody responses. Similarly,
442 certain S mutations have been reported to cause escape from T cell recognition (*10*, *11*).
443 Consequently, immunodominant regions of the viral proteome may be expected to have high
444 mutation rates to escape from T cell immunosurveillance. However, we found that non-S
445 immunogenic epitopes were enriched in the less diverse regions of the SARS-CoV-2 proteome.
446 Our data suggest that the high conservation and lack of immune editing in immunogenic regions
447 is due to their positions in functionally and structurally important regions of the proteome, thus
448 subject to evolutionary pressure restricting mutations and maintaining viral fitness. Another
449 explanation is that MHC polymorphisms and diversity across populations provide an obstacle for
450 viral immune evasion from T cell immunity.

451 Together, our findings provide a comprehensive dataset of functional T cell responses across
452 MHC-typed study cohorts, which will help to greatly inform three ongoing research efforts in
453 this field: i) technologies identifying SARS-CoV-2 epitopes and cognate TCRs to track antigen-
454 specific T cell immunity (*63*, *64*) ii) T cell-based diagnostics and correlates of protection to
455 complement antibody-based metrics and iii) design of next-generation coronavirus vaccines.

456 MATERIALS AND METHODS

457 Study Design

458 This study aimed to characterize SARS-CoV-2-specific T cell immunity elicited by vaccination
459 or natural infection. We evaluated the kinetics of Spike-specific adaptive immunity induction
460 following mRNA-based SARS-CoV-2 vaccination utilizing blood samples collected before
461 vaccination, 14 days after 1st dose, 7 days after 2nd dose, and 14 days after 2nd dose. We also
462 evaluated how Spike mutations present in variants of concern impact T cell recognition.
463 Furthermore, we investigated the breadth and magnitude of T cell responses against SARS-CoV-
464 2 in unvaccinated, convalescent individuals. We mapped immunogenic T cell epitopes within
465 non-spike proteins and evaluated their distributions across the viral proteome in the context of
466 mutational diversity. SARS-CoV-2-specific antibody responses were measured in binding assays
467 and pseudovirus neutralization assays. SARS-CoV-2-specific T cell responses were measured in
468 *ex vivo* ELISPOT and antigen-specific T cell expansion assays utilizing overlapping peptides,
469 15mers with 5 amino acids offset. Sample sizes and statistical tests used were indicated in the
470 figure legends and/or relevant method sections.

471 **Study Population**

472 The study populations included individuals receiving initial two doses of mRNA-based COVID-
473 19 vaccines, and two cohorts of unvaccinated, convalescent patients. The collection, processing
474 and banking of vaccinated donor blood specimens were carried out by the Mount Sinai Human
475 Immune Monitoring Core (HIMC). The sample collection for the first convalescent patient
476 cohort (CPC; utilized in fig. 1E and 2B) was performed by the Cancer Immunotherapy Clinical
477 Trials Team at Tisch Cancer Institute at Mount Sinai. PBMCs were isolated by density gradient
478 centrifugation using Ficoll-Paque™ Plus (GE Healthcare) and cryopreserved in human serum
479 containing 10% DMSO. The collection, processing and banking of blood samples from second
480 convalescent patient cohort (ATLAS; utilized in fig. 3 and 4) were performed by the Mount Sinai
481 Convalescent Plasma Donor Program. The use of patient-derived specimens was approved by the
482 Institutional Review Boards at Mount Sinai and all patients provided written informed consent
483 before the initiation of any study procedures. For ATLAS cohort a numerical COVID-19 disease
484 severity scoring system was adapted from Moderbacher *et al.* (*14*) and indicated in the table S1
485 together with a summary of clinical demographics and sequence-based MHC-I/II genotyping
486 (SBT) results (Histogenetics) for each subject. Although the infecting viral strains were not
487 validated by sequencing, given the circulating SARS-CoV-2 strains during the time of infection,
488 the convalescent patients in both cohorts are expected to have been infected with Wuhan-1.
489 Historical healthy donor specimens used as control were procured from New York Blood Center
490 as leukopak prior to 2019 and PBMCs were isolated by density gradient centrifugation using
491 Ficoll-Paque™ Plus (GE Healthcare). PBMCs were cryopreserved in human serum containing
492 10% DMSO.

493 **Peptide Selection and Synthesis**

494 Custom libraries of overlapping peptides (OLPs) were chemically synthesized by GenScript and
495 each peptide had at least 85% purity as determined by high-performance liquid chromatography.
496 OLPs were typically 15mers and overlapped by 10 amino acids spanning the entire spike and
497 nucleocapsid and selected regions from other SARS-CoV-2 proteins (Fig. 3A). To select regions
498 in ORF1ab, ORF3a, envelope, membrane, ORF6, ORF7a, ORF7b, ORF8 and ORF10 proteins,
499 we utilized a list of epitopes predicted to be immunogenic, published by Campbell *et al.* (*43*),
500 and prioritized regions enriched in peptide-MHC (pMHC) interactions across different races (*44*)
501 using IEDB's population coverage tool (http://tools.iedb.org/population/help/#by_ethnicity).
502 This selection was supplemented with additional epitopes that had a predicted IC50 value less
503 than 500 nM as determined using NetMHC 3.4. A complete list of all synthesized peptides and
504 peptide pooling strategies are included in table S2.

505 **Pseudovirus Neutralization Assay**

506 The neutralization capacity of patient plasma was assessed using a pseudotype particle (pp)
507 infection system. Vesicular Stomatitis Virus (VSV) pseudotype particles were generated with a
508 co-transfection strategy. This pseudovirus system was generated as described previously (*65*). A
509 VSV[Rluc]- ΔG-G construct, which encodes the viral core genes with a renilla luciferase
510 reporter substituted in place of the VSV entry surface glycoprotein gene, was used to infect 293T
511 cells. These cells were also transfected with a plasmid encoding the full-length SARS-CoV-2 S
512 from the codon-optimized Wuhan-Hu-1 isolate (NCBI accession no. NC_045512.2) and
513 containing the D614G mutation. The resulting transfection product was a particle that
514 encapsulated the viral RNA genome encoding the luciferase reporter and incorporated plasma-

515 membrane expressed SARS-CoV-2 spike proteins. CoV2pp (VSVΔG-Rluc bearing the SARS-
516 CoV-2 spike glycoprotein) at a concentration of ~400 TCID50/mL was incubated with diluted,
517 heat-inactivated patient plasma and then used to infect ACE2+ TMPRSS2+ 293T target cells.
518 Following cell lysis (Promega), relative luminescence units (RLUs) were measured with a
519 Cytation 3 reader. RLU values were normalized to those derived from cells infected with
520 CoV2pp incubated in the absence of plasma. VSV-Gpp (VSVΔG-Rluc bearing the VSV-G entry
521 glycoprotein) infection was used as a positive infection control and BALDpp (VSVΔG-Rluc
522 bearing no protein) was used as a negative infection control. This assay has been validated in our
523 laboratory using COVID-19 convalescent patient plasma and negative control plasma collected
524 from healthy donors prior to the pandemic. The 4-point nonlinear regression curves were used to
525 calculate ID50 values (GraphPad Prism v9) for each convalescent patient and vaccinated
526 individual.

527 **Luminex Antibody Binding Assay**

528 To detect antibody reactivity to the SARS-CoV-2 receptor binding domain (RBD) in individuals
529 vaccinated with COVID-19 mRNA vaccines, we used a Luminex Binding Assay described
530 previously (*66*). Briefly, SARS-CoV-2 mutant and wild-type (Wuhan-1 isolate) RBDs were
531 covalently coupled to a uniquely labeled fluorochrome carboxylated xMAP bead set (Luminex)
532 at 4.0 µg protein/million beads using a 2-step carbodiimide reaction with the xMAP Antibody
533 Coupling Kit (Luminex). The coupled beads were pelleted, resuspended at 5×10^5 beads/mL in
534 storage buffer (PBS containing 0.1% bovine serum albumin (BSA), 0.02% Tween-20, and 0.05%
535 sodium azide, pH 7.4), and stored at –80°C. The beads needed for a single run (2500 beads/well
536 × number of wells) were diluted in assay buffer (PBS containing 0.1% BSA, 0.02% Tween-20)
537 to a volume that delivered 2500 beads to each well in an aliquot of 50 µL/well. Serum/plasma
538 was diluted in PBS, added as 50 µL/well to the wells containing the beads, and incubated at
539 room temperature (RT) for 1 hour on a plate shaker at 600 rpm. After 2 washes with assay
540 buffer, 100 µL/well of biotinylated antihuman total immunoglobulin (Abcam) at 2 µg/mL was
541 added and incubated for 30 minutes at RT on a plate shaker. After 2 washes, 100 µL/well of
542 streptavidin-PE (BioLegend) at 1 µg/mL was added and followed by a 30-minute incubation at
543 RT on a plate shaker. After 2 additional washes, 100 µL of assay buffer/well was added and put
544 on a shaker to resuspend the beads. The plate was read with a Luminex Flexmap 3D instrument.

545 **Enzyme-Linked Immunosorbent Assay (ELISA)**

546 SARS-CoV-2 Spike-specific IgG antibody reactivity was assessed by ELISA. Briefly, 1 µg/mL
547 of recombinant protein (RBD or Spike ectodomain) or 100 uM peptide was coated onto Nunc
548 Maxisorp high-protein binding plates in PBS, overnight at 4°C. Plates were washed 6 times with
549 washing solution (3 times with 1xPBS with 0.2% Tween20 and 3 times with 1X PBS) and
550 incubated with blocking buffer (1xPBS with 5% dried milk powder) for 2 hours. After blocking,
551 the plates were washed similarly, and patient serum was diluted with blocking buffer and left to
552 incubate at room temperature (RT) for 2 hours. The plates were again washed prior to the
553 addition of anti-human IgG conjugated to alkaline phosphatase (AP) at a 1:3500 dilution in
554 blocking buffer. After 1 hour incubation at RT, the plates were developed with 0.6 mg/mL of
555 substrate (AttoPhos, Promega) for 30 minutes and the developing reaction was stopped with 3M
556 NaOH. The fluorescence was measured at 450nm (excitation)/555nm (emission) wavelength
557 with an ELISA microplate reader (BioTek Synergy). The anti-SARS-COV-2 Spike monoclonal
558 antibody CR3022 (Abcam) was used as a positive control for recombinant protein ELISAs.

559 **T Cell Immunogenicity Assays**

560 SARS-CoV-2-specific T cell immunogenicity was evaluated by two methods: enzyme-linked
561 immunosorbent spot (ELISpot) assays to measure *ex vivo* T cell responses and/or T cell
562 expansion assays to robustly analyze peptide immunogenicity and elicited polyfunctional
563 responses by CD8+ and CD4+ T cell subsets. For both assays, cryopreserved PBMCs were
564 quickly thawed in 37˚C water bath and transferred into RPMI medium (Thermo Fisher
565 Scientific) containing DNase I (Sigma-Aldrich) at a final concentration of 2 U/mL, spun down
566 and resuspended in media prior to being used in an assay described below.

567 For ELISpot assays, thawed PBMCs were resuspended in CTL-Test medium (ImmunoSpot)
568 supplemented with GlutaMAX (Gibco). Cells were seeded at $2.5x10^5$ cells per well in at least
569 duplicates in mixed cellular ester membrane plates (Millipore) which were previously (1-7 days)
570 coated with 4 μg/mL of anti-IFN-γ antibody (clone 1-D1k, Mabtech) and blocked by incubating
571 at least for 1 hour with RPMI medium containing 10% human serum. Cells were then stimulated
572 with 1 μM of test peptides (custom peptide synthesis, GenScript) or control reagents as indicated
573 in each relevant figure legend and costimulatory antibodies anti-CD28 (BD Biosciences) and
574 anti-CD49d (BD Biosciences) added at a final concentration of 0.5 mg/mL. After 24 hours of
575 incubation at 37˚C, plates were processed for IFN-γ detection. Plates were first incubated with
576 0.2 mg/mL of biotinylated anti-IFN-γ antibody (clone 7-B6-1 by Mabtech) for 2 hours at 37˚C,
577 then 1 hour at room temperature (RT) with 0.75 U/mL streptavidin-AP conjugate (Roche) and
578 lastly with the SigmaFast BCIP/NBT substrate for 15 minutes at RT. In between each step, plates
579 were washed 6 times with PBS containing 0.05% Tween-20 and 3 times with purified water.
580 Plates were scanned and analyzed using ImmunoSpot software.

581 For T cell expansion assays, a previously published protocol was utilized (*35*). Briefly, PBMCs
582 were resuspended in X-VIVO 15 medium (Lonza) supplemented with cytokines promoting
583 dendritic cell (DC) differentiation, GM-CSF (SANOFI, 1000 IU/mL), IL-4 (R&D Systems, 500
584 IU/mL) and Flt3L (R&D Systems, 50 ng/mL). Cells were seeded at 10^5 cells per well in U-
585 bottom 96-well plates and cultured for 24 hours before being stimulated with control reagents or
586 pooled test peptides (custom peptide synthesis, GenScript), where each peptide was at a final
587 concentration of 1 μM, together with adjuvants promoting DC maturation, LPS (Invivogen, 0.1
588 mg/mL), R848 (Invivogen, 10 mM) and IL-1β (R&D Systems 10 ng/mL), in X-VIVO 15
589 medium. Starting 24 hours after stimulation, cells were fed every 2-3 days with cytokines
590 supporting T cell expansion, IL-2 (R&D Systems, 10 IU/mL), IL-7 (R&D Systems, 10 ng/mL)
591 and IL-15 (Peprotech, 10 ng/mL) in complete RPMI media (GIBCO) containing 10% human
592 serum (R10). After 10 days of culture, cells were harvested, pooled within groups, washed,
593 resuspended in R10 and seeded in equal numbers into U-bottom 96-well plates. Expanded T cells
594 were then re-stimulated with control reagents or 1 μM of test peptides, either pooled or
595 individual, together with 0.5 mg/mL of costimulatory antibodies, anti-CD28 (BD Biosciences)
596 and anti-CD49d (BD Biosciences), and protein transport inhibitors BD GolgiStopTM, containing
597 monensin and BD GolgiPlugTM, containing brefeldin A. After 8 hours of incubation at 37˚C,
598 cells were processed for intracellular staining for flow cytometry using BD Cytofix/Cytoperm™
599 reagents according to manufacturer's protocol. The following antibodies were used: for surface
600 staining CD3 (clone SK7, FITC), CD4 (clone RPA-T4, BV785) and CD8a (clone RPA-T8,
601 APC) and for intracellular staining IFN-g (clone B27, PE), TNF-a (clone Mab11, PE/Cy7) and
602 IL-2 (clone MQ1-17H12, BV605). All antibodies were purchased via BioLegend. LIVE/DEAD

Fixable Blue Dead Cell Stain Kit by Thermo Fischer Scientific was used for live and dead cell discrimination. Data was acquired using the BD Fortessa and FlowJo V10 was used for analysis.

For both assays, DMSO (Sigma-Aldrich) was used at the equal volume of the test peptides and served as the vehicle/negative control. Unless noted otherwise all T cell immunogenicity data was shown after background normalization by subtracting the average of DMSO values (2 to 6 replicates per test group) from the corresponding spot numbers or cytokine+ cell frequencies following peptide stimulation. For peptide selection in Fig. 4 and 5, a peptide was considered immunogenic, a "hit", if in at least one patient, tested the % of reactive cells was greater than the paired DMSO % average plus 3 times the standard deviation of all DMSO values across the population (>DMSO + 3SD).

Modeling of T cell recognition and Model Selection

To model T cell responses (m) against spike mutations found in Alpha, Beta and Gamma variants (MT) or the corresponding WT (Wuhan-1) sequences in fig. 2C, we used a data likelihood approach. The complete set of experimental data measurements is represented as $D=\{m_p^{A_\kappa,B_\kappa}:\kappa\in\{\text{Alpha,Beta,Gamma}\};\ p\in\text{patients};\ A,B\in\{\text{WT, MT}\}\}$, where A_κ stands for the pool (WT or MT) that was used for the first stimulation, and B_κ (also WT or MT) is the pool used for restimulation.

We proposed a mechanistic model to account for $m_p^{A_\kappa,B_\kappa}$ in terms of the putative immunogenicity of peptides included in the stimulation and restimulation pools. Due to the presence of multiple sources of noise, including experimental conditions and intrinsic differences in patient-specific immune responses, we assumed the experimental measurements to be drawn from an underlying Gaussian process with standard deviation σ and patient-specific immune amplitudes, c_p. Our model prediction corresponding to $m_p^{A_\kappa,B_\kappa}$ is $c_p R(B_\kappa|H_p,A_\kappa,\Theta)$, where $R(B_\kappa|H_p,A_\kappa)$ is the T cell recognition of the restimulation pool, H_p is the patient's set of 6 MHC class I alleles, and Θ the model parameters. The likelihood of observing the experimental data, D, can be expressed with the Gaussian (log) likelihood function,

$$L\left(D\middle|\{c_p\}_{p\in\text{patients}},\Theta\right)\propto-\frac{1}{2\sigma^2}\sum_{p\in\text{patients}}\sum_{\kappa\in\{\alpha,\beta,\gamma,\ldots\}}\sum_{A_\kappa=\text{WT,MT}}\sum_{B_\nu=\text{WT,MT}}\left(c_p R(B_\kappa|H_p,A_\kappa,\Theta)-m_p^{A_\kappa,B_\kappa}\right)^2$$

By maximizing the log-likelihood function we find the model parameters, Θ, which minimize the difference between our model predictions and the experimental data. The optimal patient-specific immune factors can be derived analytically by setting the partial derivative of the log likelihood function equal to zero and solving for c_p, $c_p=\frac{\sum_\kappa\sum_{A_\kappa}\sum_{B_\kappa}m_p^{A_\kappa,B_\kappa}R(B_\kappa|H_p,A_\kappa)}{\sum_\kappa\sum_{A_\kappa}\sum_{B_\kappa}R^2(B_\kappa|H_p,A_\kappa)}$. We used the Hyperopt package to optimize the remaining model parameters (*67*).

We considered T cell response as a two-step process: 1) peptide binding and presentation by MHC class I molecules (pMHC) and 2) binding of pMHCs by specific T cell receptors (TCRs). Accordingly, we computed T cell recognition at the 9mer level before aggregating over all 9mers in the peptide pools, and over the MHC-I set for each patient. The T cell recognition score for a single 9mer, **b**, from the restimulation pool B, in isolation from other 9mers in the pool, is computed as $R(\mathbf{b},h)=p_{MHC}(\mathbf{a},h,A)p_{MHC}(\mathbf{b},h,B)\exp[-\beta d(\mathbf{a},\mathbf{b})]$ The first term on the right-hand side accounts for the strength of presentation of b by an MHC molecule of allele

h, $p_{MHC}(\mathbf{a}, h, B) \propto \frac{1}{K_d(\mathbf{a},h)}$. The second presentation term and the exponential term account for the cross-reactivity of peptides. The cross-reactivity distance, $d(\mathbf{a}, \mathbf{b})$, is a linear, sequence-based distance function derived from extensive epitope-TCR binding assays (*42*). In our equation we accounted for cross-reactivity of the corresponding WT-mutant peptides based on the assumption that 9mers present during the first stimulation (e.g., WT, **a**) which were well-presented would likely have elicited expansion of their cognate TCRs. Those expanded TCRs are then available to recognize restimulation pool 9mers (e.g.,Mutant, **b**), the binding strength depending on the extent of cross-reactivity with the first stimulation pool 9mers. The corresponding WT-mutant peptides only differed in most cases by a single amino acid; the cross-reactivity of the remaining pairs of peptides, depicted in fig. S3, are predicted to be negligible.

This basic model was extended with other components. Because all patients in this cohort were previously vaccinated, we can optionally include the vaccine as a zeroth stimulation event. This contributes another presentation term - for all spike protein 9mers, as taken from the sequence of the Wuhan-1 strain used in the original vaccine - and two additional cross-reactivity terms with the corresponding vaccine 9mer,

$$R(\mathbf{b},h)=p_{vacc}(\mathbf{a_v},h,V)p_{MHC}(\mathbf{b},h,B)p_{MHC}\times\exp\left[-\beta\left(d(\mathbf{a},\mathbf{b})+d(\mathbf{a}_{vacc},\mathbf{a})+d(\mathbf{a}_{vacc},\mathbf{b})\right)\right]$$

We can also include sequence similarity to immunogenic epitopes in the IEDB database (www.iedb.org). This approach has been previously used to define a TCR-immunogenicity score for neoantigenic peptides, attributing higher microbial peptide sequence similarity with increased TCR response (*40*, *41*). This introduces a new weight-parameter for the IEDB-similarity term, β_{IEDB},$R(\mathbf{b}|h)=p_{pres}(\mathbf{b},h,B)*p_{pres}(\mathbf{a},h,A)*\exp\left[-\beta\, d(\mathbf{a}, \mathbf{b})-\beta_{IEDB}\, d_{min}(\mathbf{b},IEDB)\right]$, where $d_{min}(\mathbf{b},IEDB)$ is the distance of peptide **b** to the closest IEDB peptide.

We also tested different epitope dominance models to account for the contribution of different 9mers in the peptide pools. The immunodominance assumption is that the most immunogenic 9mer in a pool is largely responsible for the observed T cell response. Mathematically, this can be achieved by taking the maximum over all 9mer recognition terms. The maximum likelihood modeling framework was used to directly test the immunodominance hypothesis, and to compare it to a number of other assumptions using alternate aggregating functions including "sum", where all 9mers contribute to the pool level response in proportion to their relative immunogenicity, "minimum", not expected to be biologically realistic, and the "Boltzmann operator", a parameterized smooth approximation to both max ($\alpha\to\infty$) and min ($\alpha\to-\infty$):

$$boltz_\alpha(x) = \frac{\sum_{i=1}^{n} x_i\, exp(\alpha x_i)}{\sum_{i=1}^{n} exp(\alpha x_i)}$$

We assumed that all MHC-I alleles of a patient contribute to the pool-level response in a proportional manner, necessitating the summation as the aggregating function over H_p, the set of 6 MHC-I alleles of the patient. The pool level response with 9mer-aggregating function F is therefore, $R\left(B_\kappa \middle| H_p, A_\kappa, \Theta\right)=\sum_{h\in H_p} F_{b\in B_\kappa}(R(\mathbf{b}, h))$ with model parameters $\Theta=\{\beta,\alpha\}$.

To compare the performance of these different models, we calculated the Bayesian Information Criterion (BIC), $BIC=k\log n-2\log\hat{L}$, where $\hat{L}$ is the maximum log-likelihood score of a given model, k is the number of parameters of this model and n is the data size.

683 The recognition scores, $\mathrm{R}(\mathbf{b}|h)$, were evaluated for each subsequence 9mer **b** of the 15mer
684 peptides in the pools, Alpha, Beta and Gamma with 161, 161, and 203 subsequences,
685 respectively, for each MHC allele h of the patients in the cohort. We computed the effective
686 number of peptides that can be recognized among all the 9mers in a pool, the *perplexity*, $P =$
687 $exp(H_{\mathrm{rec}})$, with entropy $H_{\mathrm{rec}} = -\sum_{\mathbf{b}\in\mathbf{pool}} R(\mathbf{b}|h)/Z \; log \; R\,(\mathbf{b}|h)/Z$ and with $Z = \sum_b R(\mathbf{b}|h)$,
688 the normalization constant assuring the entropy is computed over a probability distribution.

689 **Novelty of Immunogenic T Cell Epitopes**

690 To determine the novelty of hit peptides, minimal epitopes binding to donor MHC alleles were
691 predicted. For CD4 epitope analysis, 98 hit peptides (15mers, table S3) were included and
692 predictions were performed using NetMHCIIpan 4.0. Epitopes (9-15mer) with a predicted
693 binding affinity lower than 1000 nM were analyzed for their presence in IEDB. Similarly, for
694 CD8 epitope analysis, 38 hit peptides (15mers, table S3) were included and predictions were
695 performed using MHCflurry 2.0. Epitopes (9-11mer) with a predicted binding affinity lower than
696 500 nM were analyzed for their presence in IEDB. The determination of possible minimal
697 epitopes was MHC-dependent, but the IEDB search was MHC-independent. The database was
698 exported in August 2022. The MHC alleles of all responsive donors were included donors. A
699 15mer was considered novel if none of its predicted binders, across all patient MHC alleles, were
700 present in IEDB. A permutation test was used to assess the significance of the novelty
701 calculations for CD8+ T cell epitopes. Specifically, we generated synthetic lists of hits by
702 random sampling without replacement from the full set of observations (peptides x patients). We
703 did this 1000 times and ran the resulting lists of hits through the same pipeline to calculate the
704 fractions that are novel. Using this resampling procedure, we found that the number of novel
705 peptides we would expect to see by chance is significantly different ($p<0.001$) than what we
706 observed.

707 **Conservation of T Cell Epitopes**

708 To assess the conservation of selected immunogenic CD4+ and CD8+ T cell epitopes, the
709 percentage of GISAID sequences deposited that have an exact match to the reactive peptide were
710 measured. 100,000 random isolates from the GISAID database were subsampled and filtered to
711 those with a unique accession number, where no proteins contain an unknown amino acid (X)
712 after translation, have the full complement of all 27 proteins annotated, and were deposited
713 during 2022. This analysis excluded ORF10 since it is not found in GISAID.

714 The degree of amino acid residue sharing throughout the human coronavirus family was
715 determined by multiple sequence alignment using CLUSTAL O (1.2.4) for H-COV
716 Nucleocapsid Proteins: 229E (UniProt Accession: A0A127AU35), NL63 (UniProt Accession:
717 Q06XQ2), HKU1 (UniProt Accession: Q5MQC6), OC43 (UniProt Accession: P33469), SARS_2
718 (UniProt Accession: P0DTC9).

719 Peptide binding predictions in fig. 5C and S10 were performed using netMHCpan-4.1 and
720 netMHCIIpan-4.1 for MHC-I and MHC-II predictions, respectively. Peptide binding thresholds
721 were set to 2% for MHC-I and 10% for MHC-II (*68*).

722 **Diversity Analysis**

723 To assess mutational diversity during global circulation in the SARS-CoV-2 genome, entropy of
724 the observed amino acid frequencies was computed. The sequence and regional epidemiological

count data were combined to account for regional disproportions in sequencing efforts while computing amino acid frequencies. The sequence data used was obtained from the GISAID EpiCoV database (*46*) available until 2024-03-21. For quality control, the 3' and 5' regions of sequences were truncated and sequences that contain more than 5% ambiguous sites or have an incomplete collection date were removed. All sequences were aligned against a reference isolate from GenBank 80 (MN908947), using MAFFT v7.525 (*69*). Weekly infection rates were used for individual countries as reported from the WHO Coronavirus dashboard (*47*) (download date: 2024-03-21). Infection counts were distributed equally over all days within a given reporting week. Individual countries were grouped into coarse grained geographical regions r by their continents (North America, Europe, Asia, South America, Africa and Oceania). The sequence counts in each geographical region r and day t were computed as $N_r(t) = \sum_{i \in r} \omega(t - t_i)$, where $\omega(\tau) = \exp(-\tau^4/4\sigma^2)$ with $\sigma = 11$ days is the smoothing kernel function. Combining the sequence counts with the corresponding incidence data, $I_r(t)$, weight factors $m_r(t) = I_r(t)/N_r(t)$ measuring the incidence per sequence count in each region were recorded (*9*, *48*). Each GISAID sequence collected at region r and on date t was weighted by a factor $m_r(t)/\sum_{r,t} m_r(r)$ when computing the amino acid frequencies $x(a, i)$ on individual codon positions i. The entropy for position i was then computed as $H(i) = -\sum_a x(a, i)\, log\, x\,(a, i)$.

To determine where experimentally validated immunogenic T cell epitopes reside, the entropy value for each codon (and its corresponding residue) plotted as a bar graph was superimposed with an immunogenicity value for each individual residue across the nucleocapsid protein. Per residue immunogenicity values were calculated as the sum of the reactive T cell percentages against each peptide spanning that residue.

As an alternative approach, Shannon entropy data from Nextstrain available in August 2022 and February 2023 (https://nextstrain.org/ncov/gisaid/global) were analyzed for each codon. These entropy values were calculated using complete viral genome sequences deposited to the Global Initiative on Sharing All Influenza Data (GISAID) from around the world. This approach was not applied to analysis by April 2024 since the most recent GISAID datasets did not include codon/residue entropies.

Structural Analysis of Nucleocapsid

To assess whether immunogenic T cell epitopes contained structurally important residues, immunogenic regions were mapped onto 6YVO and 6ZWO crystal structures, RNA binding and dimerization domains of nucleocapsid, respectively. The change in protein stability (DDG) was measured by mutating each residue of the protein into alanine and comparing the energy of that structure to the wildtype. This resulted in a DDG value being assigned to every residue with positive values indicating destabilizing mutations and negative values indicating stabilizing mutations. The top ten most immunogenic peptides were isolated from each patient, and the average DDG for those peptides was determined by averaging the values of individual residues. Alanine scanning was performed using the Rosetta analysis suite (*70*). The impacts on monomer stability of the RBD were probed using the cartesian_ddg application (*71*).The interface of the dimerization domain was analyzed using the Robetta alanine scanning server of the 6ZWO crystal structure (*72*).

Statistical Analysis

Statistical tests used are indicated in the figure legends and in relevant methods sections. Briefly, statistical analyses were performed using GraphPad Prism versions 8 and 9. All error bars

769 represent mean + SD unless noted otherwise in the figure legend. To evaluate significance,
770 Wilcoxon matched-pairs test was used for paired samples and Welch's t-test was used for
771 unpaired samples. Significance was denoted by *, **, ***, ****, indicating $P < 0.05$, $P < 0.01$, P
772 < 0.001, and $P < 0.0001$, respectively.

773

774 **Supplementary Materials**

775 Figs. S1 to S13

776 Tables S1 to S3

777

778 **References and Notes**

779 1. R. M. Meganck, C. E. Edwards, M. L. Mallory, R. E. Lee, H. Dang, A. B. Bailey, J. A.
780 Wykoff, S. C. Gallant, D. R. Zhu, B. L. Yount, T. Kato, K. M. Shaffer, S. Nakano, A. M.
781 Cawley, V. Sontake, J. R. Wang, R. S. Hagan, M. B. Miller, P. R. Tata, S. H. Randell, L. V.
782 Tse, C. Ehre, K. Okuda, R. C. Boucher, R. S. Baric, SARS-CoV-2 variant of concern fitness
783 and adaptation in primary human airway epithelia. *Cell Rep.* **43**, 114076 (2024).

784 2. T. M. Wilkinson, C. K. F. Li, C. S. C. Chui, A. K. Y. Huang, M. Perkins, J. C. Liebner, R.
785 Lambkin-Williams, A. Gilbert, J. Oxford, B. Nicholas, K. J. Staples, T. Dong, D. C. Douek,
786 A. J. McMichael, X.-N. Xu, Preexisting influenza-specific CD4+ T cells correlate with
787 disease protection against influenza challenge in humans. *Nat. Med.* **18**, 274–280 (2012).

788 3. D. Stadlbauer, X. Zhu, M. McMahon, J. S. Turner, T. J. Wohlbold, A. J. Schmitz, S.
789 Strohmeier, W. Yu, R. Nachbagauer, P. A. Mudd, I. A. Wilson, A. H. Ellebedy, F.
790 Krammer, Broadly protective human antibodies that target the active site of influenza virus
791 neuraminidase. *Science* **366**, 499–504 (2019).

792 4. M. Bergwerk, T. Gonen, Y. Lustig, S. Amit, M. Lipsitch, C. Cohen, M. Mandelboim, E. G.
793 Levin, C. Rubin, V. Indenbaum, I. Tal, M. Zavitan, N. Zuckerman, A. Bar-Chaim, Y.
794 Kreiss, G. Regev-Yochay, Covid-19 Breakthrough Infections in Vaccinated Health Care
795 Workers. *N. Engl. J. Med.* **385**, 1474–1484 (2021).

796 5. W. F. Garcia-Beltran, E. C. Lam, M. G. Astudillo, D. Yang, T. E. Miller, J. Feldman, B. M.
797 Hauser, T. M. Caradonna, K. L. Clayton, A. D. Nitido, M. R. Murali, G. Alter, R. C.
798 Charles, A. Dighe, J. A. Branda, J. K. Lennerz, D. Lingwood, A. G. Schmidt, A. J. Iafrate,
799 A. B. Balazs, COVID-19-neutralizing antibodies predict disease severity and survival. *Cell*
800 **184**, 476-488.e11 (2021).

801 6. J. P. Evans, C. Zeng, C. Carlin, G. Lozanski, L. J. Saif, E. M. Oltz, R. J. Gumina, S.-L. Liu,
802 Neutralizing antibody responses elicited by SARS-CoV-2 mRNA vaccination wane over
803 time and are boosted by breakthrough infection. *Sci. Transl. Med.* **14**, eabn8057 (2022).

804 7. W. F. Garcia-Beltran, E. C. Lam, K. St Denis, A. D. Nitido, Z. H. Garcia, B. M. Hauser, J.
805 Feldman, M. N. Pavlovic, D. J. Gregory, M. C. Poznansky, A. Sigal, A. G. Schmidt, A. J.

Iafrate, V. Naranbhai, A. B. Balazs, Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell* **184**, 2372-2383.e9 (2021).

8. M. Hoffmann, N. Krüger, S. Schulz, A. Cossmann, C. Rocha, A. Kempf, I. Nehlmeier, L. Graichen, A.-S. Moldenhauer, M. S. Winkler, M. Lier, A. Dopfer-Jablonka, H.-M. Jäck, G. M. N. Behrens, S. Pöhlmann, The Omicron variant is highly resistant against antibody-mediated neutralization: Implications for control of the COVID-19 pandemic. *Cell* **185**, 447-456.e11 (2022).

9. M. Meijers, D. Ruchnewitz, J. Eberhardt, M. Łuksza, M. Lässig, Population immunity predicts evolutionary trajectories of SARS-CoV-2. *Cell* **186**, 5151-5164.e13 (2023).

10. C. Motozono, M. Toyoda, J. Zahradnik, A. Saito, H. Nasser, T. S. Tan, I. Ngare, I. Kimura, K. Uriu, Y. Kosugi, Y. Yue, R. Shimizu, J. Ito, S. Torii, A. Yonekawa, N. Shimono, Y. Nagasaki, R. Minami, T. Toya, N. Sekiya, T. Fukuhara, Y. Matsuura, G. Schreiber, Genotype to Phenotype Japan (G2P-Japan) Consortium, T. Ikeda, S. Nakagawa, T. Ueno, K. Sato, SARS-CoV-2 spike L452R variant evades cellular immunity and increases infectivity. *Cell Host Microbe* **29**, 1124-1136.e11 (2021).

11. G. Dolton, C. Rius, M. S. Hasan, A. Wall, B. Szomolay, E. Behiry, T. Whalley, J. Southgate, A. Fuller, COVID-19 Genomics UK (COG-UK) consortium, T. Morin, K. Topley, L. R. Tan, P. J. R. Goulder, O. B. Spiller, P. J. Rizkallah, L. C. Jones, T. R. Connor, A. K. Sewell, Emergence of immune escape at dominant SARS-CoV-2 killer T cell epitope. *Cell* **185**, 2936-2951.e19 (2022).

12. S. Meyer, I. Blaas, R. C. Bollineni, M. Delic-Sarac, T. T. Tran, C. Knetter, K.-Z. Dai, T. S. Madssen, J. T. Vaage, A. Gustavsen, W. Yang, L. S. H. Nissen-Meyer, K. Douvlataniotis, M. Laos, M. M. Nielsen, B. Thiede, A. Søraas, F. Lund-Johansen, E. H. Rustad, J. Olweus, Prevalent and immunodominant CD8 T cell epitopes are conserved in SARS-CoV-2 variants. *Cell Rep.* **42**, 111995 (2023).

13. A. Tarke, C. H. Coelho, Z. Zhang, J. M. Dan, E. D. Yu, N. Methot, N. I. Bloom, B. Goodwin, E. Phillips, S. Mallal, J. Sidney, G. Filaci, D. Weiskopf, R. da Silva Antunes, S. Crotty, A. Grifoni, A. Sette, SARS-CoV-2 vaccination induces immunological T cell memory able to cross-recognize variants from Alpha to Omicron. *Cell* **185**, 847-859.e11 (2022).

14. C. Rydyznski Moderbacher, S. I. Ramirez, J. M. Dan, A. Grifoni, K. M. Hastie, D. Weiskopf, S. Belanger, R. K. Abbott, C. Kim, J. Choi, Y. Kato, E. G. Crotty, C. Kim, S. A. Rawlings, J. Mateus, L. P. V. Tse, A. Frazier, R. Baric, B. Peters, J. Greenbaum, E. Ollmann Saphire, D. M. Smith, A. Sette, S. Crotty, Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell* **183**, 996-1012.e19 (2020).

15. L. Bergamaschi, F. Mescia, L. Turner, A. L. Hanson, P. Kotagiri, B. J. Dunmore, H. Ruffieux, A. De Sa, O. Huhn, M. D. Morgan, P. P. Gerber, M. R. Wills, S. Baker, F. J. Calero-Nieto, R. Doffinger, G. Dougan, A. Elmer, I. G. Goodfellow, R. K. Gupta, M. Hosmillo, K. Hunter, N. Kingston, P. J. Lehner, N. J. Matheson, J. K. Nicholson, A. M.

Petrunkina, S. Richardson, C. Saunders, J. E. D. Thaventhiran, E. J. M. Toonen, M. P. Weekes, Cambridge Institute of Therapeutic Immunology and Infectious Disease-National Institute of Health Research (CITIID-NIHR) COVID BioResource Collaboration, B. Göttgens, M. Toshner, C. Hess, J. R. Bradley, P. A. Lyons, K. G. C. Smith, Longitudinal analysis reveals that delayed bystander CD8+ T cell activation and early immune pathology distinguish severe COVID-19 from mild disease. *Immunity* **54**, 1257-1275.e8 (2021).

16. K. McMahan, J. Yu, N. B. Mercado, C. Loos, L. H. Tostanoski, A. Chandrashekar, J. Liu, L. Peter, C. Atyeo, A. Zhu, E. A. Bondzie, G. Dagotto, M. S. Gebre, C. Jacob-Dolan, Z. Li, F. Nampanya, S. Patel, L. Pessaint, A. Van Ry, K. Blade, J. Yalley-Ogunro, M. Cabus, R. Brown, A. Cook, E. Teow, H. Andersen, M. G. Lewis, D. A. Lauffenburger, G. Alter, D. H. Barouch, Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* **590**, 630–634 (2021).

17. G. Guerrera, M. Picozza, S. D'Orso, R. Placido, M. Pirronello, A. Verdiani, A. Termine, C. Fabrizio, F. Giannessi, M. Sambucci, M. P. Balice, C. Caltagirone, A. Salvia, A. Rossini, L. Battistini, G. Borsellino, BNT162b2 vaccination induces durable SARS-CoV-2-specific T cells with a stem cell memory phenotype. *Sci Immunol* **6**, eabl5344 (2021).

18. B. Israelow, T. Mao, J. Klein, E. Song, B. Menasche, S. B. Omer, A. Iwasaki, Adaptive immune determinants of viral clearance and protection in mouse models of SARS-CoV-2. *Sci Immunol* **6**, eabl4509 (2021).

19. H. Ishii, T. Nomura, H. Yamamoto, M. Nishizawa, T. T. Thu Hau, S. Harada, S. Seki, M. Nakamura-Hoshi, M. Okazaki, S. Daigen, A. Kawana-Tachikawa, N. Nagata, N. Iwata-Yoshikawa, N. Shiwa, T. Suzuki, E.-S. Park, M. Ken, T. Onodera, Y. Takahashi, K. Kusano, R. Shimazaki, Y. Suzaki, Y. Ami, T. Matano, Neutralizing-antibody-independent SARS-CoV-2 control correlated with intranasal-vaccine-induced CD8+ T cell responses. *Cell Rep Med* **3**, 100520 (2022).

20. S. A. Apostolidis, M. Kakara, M. M. Painter, R. R. Goel, D. Mathew, K. Lenzi, A. Rezk, K. R. Patterson, D. A. Espinoza, J. C. Kadri, D. M. Markowitz, C. E Markowitz, I. Mexhitaj, D. Jacobs, A. Babb, M. R. Betts, E. T. L. Prak, D. Weiskopf, A. Grifoni, K. A. Lundgreen, S. Gouma, A. Sette, P. Bates, S. E. Hensley, A. R. Greenplate, E. J. Wherry, R. Li, A. Bar-Or, Cellular and humoral immune responses following SARS-CoV-2 mRNA vaccination in patients with multiple sclerosis on anti-CD20 therapy. *Nat. Med.* **27**, 1990–2001 (2021).

21. M. P. Sormani, I. Schiavetti, M. Inglese, L. Carmisciano, A. Laroni, C. Lapucci, V. Visconti, C. Serrati, I. Gandoglia, T. Tassinari, G. Perego, G. Brichetto, P. Gazzola, A. Mannironi, M. L. Stromillo, C. Cordioli, D. Landi, M. Clerico, E. Signoriello, E. Cocco, J. Frau, M. T. Ferrò, A. Di Sapio, L. Pasquali, M. Ulivelli, F. Marinelli, M. Pizzorno, G. Callari, R. Iodice, G. Liberatore, F. Caleri, A. M. Repice, S. Cordera, M. A. Battaglia, M. Salvetti, D. Franciotta, A. Uccelli, CovaXiMS study group, Breakthrough SARS-CoV-2 infections after COVID-19 mRNA vaccination in MS patients on disease modifying therapies during the Delta and the Omicron waves in Italy. *EBioMedicine* **80**, 104042 (2022).

886 22. A. Grifoni, D. Weiskopf, S. I. Ramirez, J. Mateus, J. M. Dan, C. R. Moderbacher, S. A.
887 Rawlings, A. Sutherland, L. Premkumar, R. S. Jadi, D. Marrama, A. M. de Silva, A. Frazier,
888 A. F. Carlin, J. A. Greenbaum, B. Peters, F. Krammer, D. M. Smith, S. Crotty, A. Sette,
889 Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19
890 Disease and Unexposed Individuals. *Cell* **181**, 1489-1501.e15 (2020).

891 23. H. Kared, A. D. Redd, E. M. Bloch, T. S. Bonny, H. Sumatoh, F. Kairi, D. Carbajo, B.
892 Abel, E. W. Newell, M. P. Bettinotti, S. E. Benner, E. U. Patel, K. Littlefield, O.
893 Laeyendecker, S. Shoham, D. Sullivan, A. Casadevall, A. Pekosz, A. Nardin, M. Fehlings,
894 A. A. Tobian, T. C. Quinn, SARS-CoV-2-specific CD8+ T cell responses in convalescent
895 COVID-19 individuals. *J. Clin. Invest.* **131** (2021).

896 24. S. K. Saini, D. S. Hersby, T. Tamhane, H. R. Povlsen, S. P. Amaya Hernandez, M. Nielsen,
897 A. O. Gang, S. R. Hadrup, SARS-CoV-2 genome-wide T cell epitope mapping reveals
898 immunodominance and substantial CD8+ T cell activation in COVID-19 patients. *Sci*
899 *Immunol* **6** (2021).

900 25. T. Sekine, A. Perez-Potti, O. Rivera-Ballesteros, K. Strålin, J.-B. Gorin, A. Olsson, S.
901 Llewellyn-Lacey, H. Kamal, G. Bogdanovic, S. Muschiol, D. J. Wullimann, T. Kammann,
902 J. Emgård, T. Parrot, E. Folkesson, Karolinska COVID-19 Study Group, O. Rooyackers, L.
903 I. Eriksson, J.-I. Henter, A. Sönnerborg, T. Allander, J. Albert, M. Nielsen, J. Klingström, S.
904 Gredmark-Russ, N. K. Björkström, J. K. Sandberg, D. A. Price, H.-G. Ljunggren, S.
905 Aleman, M. Buggert, Robust T Cell Immunity in Convalescent Individuals with
906 Asymptomatic or Mild COVID-19. *Cell* **183**, 158-168.e14 (2020).

907 26. J. Mateus, A. Grifoni, A. Tarke, J. Sidney, S. I. Ramirez, J. M. Dan, Z. C. Burger, S. A.
908 Rawlings, D. M. Smith, E. Phillips, S. Mallal, M. Lammers, P. Rubiro, L. Quiambao, A.
909 Sutherland, E. D. Yu, R. da Silva Antunes, J. Greenbaum, A. Frazier, A. J. Markmann, L.
910 Premkumar, A. de Silva, B. Peters, S. Crotty, A. Sette, D. Weiskopf, Selective and cross-
911 reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* **370**, 89–94 (2020).

912 27. N. Le Bert, A. T. Tan, K. Kunasegaran, C. Y. L. Tham, M. Hafezi, A. Chia, M. H. Y. Chng,
913 M. Lin, N. Tan, M. Linster, W. N. Chia, M. I.-C. Chen, L.-F. Wang, E. E. Ooi, S.
914 Kalimuddin, P. A. Tambyah, J. G.-H. Low, Y.-J. Tan, A. Bertoletti, SARS-CoV-2-specific
915 T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**,
916 457–462 (2020).

917 28. J. Braun, L. Loyal, M. Frentsch, D. Wendisch, P. Georg, F. Kurth, S. Hippenstiel, M.
918 Dingeldey, B. Kruse, F. Fauchere, E. Baysal, M. Mangold, L. Henze, R. Lauster, M. A.
919 Mall, K. Beyer, J. Röhmel, S. Voigt, J. Schmitz, S. Miltenyi, I. Demuth, M. A. Müller, A.
920 Hocke, M. Witzenrath, N. Suttorp, F. Kern, U. Reimer, H. Wenschuh, C. Drosten, V. M.
921 Corman, C. Giesecke-Thiel, L. E. Sander, A. Thiel, SARS-CoV-2-reactive T cells in
922 healthy donors and patients with COVID-19. *Nature* **587**, 270–274 (2020).

923 29. V. Cicaloni, F. Costanti, A. Pasqui, M. Bianchini, N. Niccolai, P. Bongini, A
924 Bioinformatics Approach to Investigate Structural and Non-Structural Proteins in Human
925 Coronaviruses. *Front. Genet.* **13**, 891418 (2022).

926 30. K. Tamminen, M. Salminen, V. Blazevic, Seroprevalence and SARS-CoV-2 cross-
927 reactivity of endemic coronavirus OC43 and 229E antibodies in Finnish children and adults.
928 *Clin. Immunol.* **229**, 108782 (2021).

929 31. E. N. Mulabbi, R. Tweyongyere, F. Wabwire-Mangen, E. Mworozi, J. Koehlerb, H.
930 Kibuuka, M. Millard, B. Erima, T. Tugume, U. Q. Aquino, D. K. Byarugaba,
931 Seroprevalence of human coronaviruses among patients visiting hospital-based sentinel
932 sites in Uganda. *BMC Infect. Dis.* **21**, 585 (2021).

933 32. L. Swadling, M. O. Diniz, N. M. Schmidt, O. E. Amin, A. Chandran, E. Shaw, C. Pade, J.
934 M. Gibbons, N. Le Bert, A. T. Tan, A. Jeffery-Smith, C. C. S. Tan, C. Y. L. Tham, S.
935 Kucykowicz, G. Aidoo-Micah, J. Rosenheim, J. Davies, M. Johnson, M. P. Jensen, G. Joy,
936 L. E. McCoy, A. M. Valdes, B. M. Chain, D. Goldblatt, D. M. Altmann, R. J. Boyton, C.
937 Manisty, T. A. Treibel, J. C. Moon, COVIDsortium Investigators, L. van Dorp, F. Balloux,
938 Á. McKnight, M. Noursadeghi, A. Bertoletti, M. K. Maini, Pre-existing polymerase-specific
939 T cells expand in abortive seronegative SARS-CoV-2. *Nature* **601**, 110–117 (2022).

940 33. V. Fumagalli, M. Ravà, D. Marotta, P. Di Lucia, E. B. Bono, L. Giustini, F. De Leo, M.
941 Casalgrandi, E. Monteleone, V. Mouro, C. Malpighi, C. Perucchini, M. Grillo, S. De Palma,
942 L. Donnici, S. Marchese, M. Conti, H. Muramatsu, S. Perlman, N. Pardi, M. Kuka, R. De
943 Francesco, M. E. Bianchi, L. G. Guidotti, M. Iannacone, Antibody-independent protection
944 against heterologous SARS-CoV-2 challenge conferred by prior infection or vaccination.
945 *Nat. Immunol.* **25**, 633–643 (2024).

946 34. Z. Zhang, J. Mateus, C. H. Coelho, J. M. Dan, C. R. Moderbacher, R. I. Gálvez, F. H.
947 Cortes, A. Grifoni, A. Tarke, J. Chang, E. A. Escarrega, C. Kim, B. Goodwin, N. I. Bloom,
948 A. Frazier, D. Weiskopf, A. Sette, S. Crotty, Humoral and cellular immune memory to four
949 COVID-19 vaccines. *Cell* **185**, 2434-2451.e17 (2022).

950 35. C. Cimen Bozkus, A. B. Blazquez, T. Enokida, N. Bhardwaj, A T-cell-based
951 immunogenicity protocol for evaluating human antigen-specific responses. *STAR Protoc* **2**,
952 100758 (2021).

953 36. A. Muik, A.-K. Wallisch, B. Sänger, K. A. Swanson, J. Mühl, W. Chen, H. Cai, D. Maurus,
954 R. Sarkar, Ö. Türeci, P. R. Dormitzer, U. Şahin, Neutralization of SARS-CoV-2 lineage
955 B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. *Science* **371**, 1152–1153
956 (2021).

957 37. J. Newman, N. Thakur, T. P. Peacock, D. Bialy, A. M. E. Elrefaey, C. Bogaardt, D. L.
958 Horton, S. Ho, T. Kankeyan, C. Carr, K. Hoschler, W. S. Barclay, G. Amirthalingam, K. E.
959 Brown, B. Charleston, D. Bailey, Neutralizing antibody activity against 21 SARS-CoV-2
960 variants in older adults vaccinated with BNT162b2. *Nat Microbiol* **7**, 1180–1188 (2022).

961 38. A. Tarke, J. Sidney, C. K. Kidd, J. M. Dan, S. I. Ramirez, E. D. Yu, J. Mateus, R. da Silva
962 Antunes, E. Moore, P. Rubiro, N. Methot, E. Phillips, S. Mallal, A. Frazier, S. A. Rawlings,
963 J. A. Greenbaum, B. Peters, D. M. Smith, S. Crotty, D. Weiskopf, A. Grifoni, A. Sette,
964 Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-
965 CoV-2 epitopes in COVID-19 cases. *Cell Rep Med* **2**, 100204 (2021).

39. B. Peters, M. Nielsen, A. Sette, T Cell Epitope Predictions. *Annu. Rev. Immunol.* **38**, 123–145 (2020).

40. M. Łuksza, N. Riaz, V. Makarov, V. P. Balachandran, M. D. Hellmann, A. Solovyov, N. A. Rizvi, T. Merghoub, A. J. Levine, T. A. Chan, J. D. Wolchok, B. D. Greenbaum, A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517–520 (2017).

41. L. A. Rojas, Z. Sethna, K. C. Soares, C. Olcese, N. Pang, E. Patterson, J. Lihm, N. Ceglia, P. Guasp, A. Chu, R. Yu, A. K. Chandra, T. Waters, J. Ruan, M. Amisaki, A. Zebboudj, Z. Odgerel, G. Payne, E. Derhovanessian, F. Müller, I. Rhee, M. Yadav, A. Dobrin, M. Sadelain, M. Łuksza, N. Cohen, L. Tang, O. Basturk, M. Gönen, S. Katz, R. K. Do, A. S. Epstein, P. Momtaz, W. Park, R. Sugarman, A. M. Varghese, E. Won, A. Desai, A. C. Wei, M. I. D'Angelica, T. P. Kingham, I. Mellman, T. Merghoub, J. D. Wolchok, U. Sahin, Ö. Türeci, B. D. Greenbaum, W. R. Jarnagin, J. Drebin, E. M. O'Reilly, V. P. Balachandran, Personalized RNA neoantigen vaccines stimulate T cells in pancreatic cancer. *Nature* **618**, 144–150 (2023).

42. M. Łuksza, Z. M. Sethna, L. A. Rojas, J. Lihm, B. Bravi, Y. Elhanati, K. Soares, M. Amisaki, A. Dobrin, D. Hoyos, P. Guasp, A. Zebboudj, R. Yu, A. K. Chandra, T. Waters, Z. Odgerel, J. Leung, R. Kappagantula, A. Makohon-Moore, A. Johns, A. Gill, M. Gigoux, J. Wolchok, T. Merghoub, M. Sadelain, E. Patterson, R. Monasson, T. Mora, A. M. Walczak, S. Cocco, C. Iacobuzio-Donahue, B. D. Greenbaum, V. P. Balachandran, Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature* **606**, 389–395 (2022).

43. Katie M. Campbell, Gabriela Steiner, Daniel K. Wells, Antoni Ribas, Anusha Kalbasi, Prioritization of SARS-CoV-2 epitopes using a pan-HLA and global population inference approach, *BioRxiv* (2020). https://doi.org/10.1101/2020.03.30.016931.

44. V. Roudko, C. C. Bozkus, T. Orfanelli, C. B. McClain, C. Carr, T. O'Donnell, L. Chakraborty, R. Samstein, K.-L. Huang, S. V. Blank, B. Greenbaum, N. Bhardwaj, Shared Immunogenic Poly-Epitope Frameshift Mutations in Microsatellite Unstable Tumors. *Cell* **183**, 1634-1649.e17 (2020).

45. M. H. Abbasian, M. Mahmanzar, K. Rahimian, B. Mahdavi, S. Tokhanbigli, B. Moradi, M. M. Sisakht, Y. Deng, Global landscape of SARS-CoV-2 mutations and conserved regions. *J. Transl. Med.* **21**, 152 (2023).

46. S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **1**, 33–46 (2017).

47. World Health Organization 2023 data.who.int, WHO Coronavirus (COVID-19) dashboard > About [Dashboard]. https://data.who.int/dashboards/covid19/about.

48. Matthijs Meijers, Denis Ruchnewitz, Jan Eberhardt, Malancha Karmakar, Marta Łuksza, Michael Lässig, Concepts and methods for predicting viral evolution, *BioRxiv* (2024). https://doi.org/10.1101/2024.03.19.585703.

1005 49. G. D. Gaiha, E. J. Rossin, J. Urbach, C. Landeros, D. R. Collins, C. Nwonu, I. Muzhingi,
1006 M. N. Anahtar, O. M. Waring, A. Piechocka-Trocha, M. Waring, D. P. Worrall, M. S.
1007 Ghebremichael, R. M. Newman, K. A. Power, T. M. Allen, J. Chodosh, B. D. Walker,
1008 Structural topology defines protective CD8+ T cell epitopes in the HIV proteome. *Science*
1009 **364**, 480–484 (2019).

1010 50. D. C. Dinesh, D. Chalupska, J. Silhan, E. Koutna, R. Nencka, V. Veverka, E. Boura,
1011 Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein.
1012 *PLoS Pathog.* **16**, e1009100 (2020).

1013 51. S. Kang, M. Yang, Z. Hong, L. Zhang, Z. Huang, X. Chen, S. He, Z. Zhou, Z. Zhou, Q.
1014 Chen, Y. Yan, C. Zhang, H. Shan, S. Chen, Crystal structure of SARS-CoV-2 nucleocapsid
1015 protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm Sin*
1016 *B* **10**, 1228–1238 (2020).

1017 52. L. Zinzula, J. Basquin, S. Bohn, F. Beck, S. Klumpe, G. Pfeifer, I. Nagy, A. Bracher, F. U.
1018 Hartl, W. Baumeister, High-resolution structure and biophysical characterization of the
1019 nucleocapsid phosphoprotein dimerization domain from the Covid-19 severe acute
1020 respiratory syndrome coronavirus 2. *Biochem. Biophys. Res. Commun.* **538**, 54–62 (2021).

1021 53. A. P. Ferretti, T. Kula, Y. Wang, D. M. V. Nguyen, A. Weinheimer, G. S. Dunlap, Q. Xu,
1022 N. Nabilsi, C. R. Perullo, A. W. Cristofaro, H. J. Whitton, A. Virbasius, K. J. Olivier Jr, L.
1023 R. Buckner, A. T. Alistar, E. D. Whitman, S. A. Bertino, S. Chattopadhyay, G. MacBeath,
1024 Unbiased Screens Show CD8+ T Cells of COVID-19 Patients Recognize Shared Epitopes
1025 in SARS-CoV-2 that Largely Reside outside the Spike Protein. *Immunity* **53**, 1095-1107.e3
1026 (2020).

1027 54. A. T. Tan, M. Linster, C. W. Tan, N. Le Bert, W. N. Chia, K. Kunasegaran, Y. Zhuang, C.
1028 Y. L. Tham, A. Chia, G. J. D. Smith, B. Young, S. Kalimuddin, J. G. H. Low, D. Lye, L.-F.
1029 Wang, A. Bertoletti, Early induction of functional SARS-CoV-2-specific T cells associates
1030 with rapid viral clearance and mild disease in COVID-19 patients. *Cell Rep.* **34**, 108728
1031 (2021).

1032 55. N. K. Routhu, S. Gangadhara, L. Lai, M. E. Davis-Gardner, K. Floyd, A. Shiferaw, Y. C.
1033 Bartsch, S. Fischinger, G. Khoury, S. A. Rahman, S. D. Stampfer, A. Schäfer, S. M. Jean,
1034 C. Wallace, R. L. Stammen, J. Wood, C. Joyce, T. Nagy, M. S. Parsons, L. Gralinski, P. A.
1035 Kozlowski, G. Alter, M. S. Suthar, R. R. Amara, A modified vaccinia Ankara vaccine
1036 expressing spike and nucleocapsid protects rhesus macaques against SARS-CoV-2 Delta
1037 infection. *Sci Immunol* **7**, eabo0226 (2022).

1038 56. S. Afkhami, M. R. D'Agostino, A. Zhang, H. D. Stacey, A. Marzok, A. Kang, R. Singh, J.
1039 Bavananthasivam, G. Ye, X. Luo, F. Wang, J. C. Ang, A. Zganiacz, U. Sankar, N. Kazhdan,
1040 J. F. E. Koenig, A. Phelps, S. F. Gameiro, S. Tang, M. Jordana, Y. Wan, K. L. Mossman,
1041 M. Jeyanathan, A. Gillgrass, M. F. C. Medina, F. Smaill, B. D. Lichty, M. S. Miller, Z.
1042 Xing, Respiratory mucosal delivery of next-generation COVID-19 vaccine provides robust
1043 protection against both ancestral and variant strains of SARS-CoV-2. *Cell* **185**, 896-915.e19
1044 (2022).

1045 57. R. L. Hajnik, J. A. Plante, Y. Liang, M.-G. Alameh, J. Tang, S. R. Bonam, C. Zhong, A.
1046 Adam, D. Scharton, G. H. Rafael, Y. Liu, N. C. Hazell, J. Sun, L. Soong, P.-Y. Shi, T.
1047 Wang, D. H. Walker, J. Sun, D. Weissman, S. C. Weaver, K. S. Plante, H. Hu, Dual spike
1048 and nucleocapsid mRNA vaccination confer protection against SARS-CoV-2 Omicron and
1049 Delta variants in preclinical models. *Sci. Transl. Med.* **14**, eabq1945 (2022).

1050 58. C. M. Arieta, Y. J. Xie, D. A. Rothenberg, H. Diao, D. Harjanto, S. Meda, K. Marquart, B.
1051 Koenitzer, T. E. Sciuto, A. Lobo, A. Zuiani, S. A. Krumm, C. I. Cadima Couto, S. Hein, A.
1052 P. Heinen, T. Ziegenhals, Y. Liu-Lupo, A. B. Vogel, J. R. Srouji, S. Fesser, K. Thanki, K.
1053 Walzer, T. A. Addona, Ö. Türeci, U. Şahin, R. B. Gaynor, A. Poran, The T-cell-directed
1054 vaccine BNT162b4 encoding conserved non-spike antigens protects animals from severe
1055 SARS-CoV-2 infection. *Cell* **186**, 2392-2409.e21 (2023).

1056 59. P. G. Thomas, S. A. Brown, R. Keating, W. Yue, M. Y. Morris, J. So, R. J. Webby, P. C.
1057 Doherty, Hidden epitopes emerge in secondary influenza virus-specific CD8+ T cell
1058 responses. *J. Immunol.* **178**, 3091–3098 (2007).

1059 60. A. R. Rappaport, C. Kyi, M. Lane, M. G. Hart, M. L. Johnson, B. S. Henick, C.-Y. Liao, A.
1060 Mahipal, A. Shergill, A. I. Spira, J. W. Goldman, C. D. Scallan, D. Schenk, C. D. Palmer,
1061 M. J. Davis, S. Kounlavouth, L. Kemp, A. Yang, Y. J. Li, M. Likes, A. Shen, G. R.
1062 Boucher, M. Egorova, R. L. Veres, J. A. Espinosa, J. R. Jaroslavsky, L. D. Kraemer Tardif,
1063 L. Acrebuche, C. Puccia, L. Sousa, R. Zhou, K. Bae, J. R. Hecht, D. P. Carbone, B.
1064 Johnson, A. Allen, A. R. Ferguson, K. Jooss, A shared neoantigen vaccine combined with
1065 immune checkpoint blockade for advanced metastatic solid tumors: phase 1 trial interim
1066 results. *Nat. Med.* **30**, 1013–1022 (2024).

1067 61. W. Zhong, P. A. Reche, C.-C. Lai, B. Reinhold, E. L. Reinherz, Genome-wide
1068 characterization of a viral cytotoxic T lymphocyte epitope repertoire. *J. Biol. Chem.* **278**,
1069 45135–45144 (2003).

1070 62. V. Oberhardt, H. Luxenburger, J. Kemming, I. Schulien, K. Ciminski, S. Giese, B.
1071 Csernalabics, J. Lang-Meli, I. Janowska, J. Staniek, K. Wild, K. Basho, M. S. Marinescu, J.
1072 Fuchs, F. Topfstedt, A. Janda, O. Sogukpinar, H. Hilger, K. Stete, F. Emmerich, B.
1073 Bengsch, C. F. Waller, S. Rieg, Sagar, T. Boettler, K. Zoldan, G. Kochs, M. Schwemmle,
1074 M. Rizzi, R. Thimme, C. Neumann-Haefelin, M. Hofmann, Rapid and stable mobilization
1075 of CD8+ T cells by SARS-CoV-2 mRNA vaccine. *Nature* **597**, 268–273 (2021).

1076 63. M. V. Pogorelyy, E. Rosati, A. A. Minervina, R. C. Mettelman, A. Scheffold, A. Franke, P.
1077 Bacher, P. G. Thomas, Resolving SARS-CoV-2 CD4+ T cell specificity via reverse epitope
1078 discovery. *Cell Rep Med* **3**, 100697 (2022).

1079 64. M. Stražar, J. Park, J. G. Abelin, H. B. Taylor, T. K. Pedersen, D. R. Plichta, E. M. Brown,
1080 B. Eraslan, Y.-M. Hung, K. Ortiz, K. R. Clauser, S. A. Carr, R. J. Xavier, D. B. Graham,
1081 HLA-II immunopeptidome profiling and deep learning reveal features of antigenicity to
1082 inform antigen discovery. *Immunity* **56**, 1681-1698.e13 (2023).

1083 65. K. Y. Oguntuyo, C. S. Stevens, C. T. Hung, S. Ikegame, J. A. Acklin, S. S. Kowdle, J. C.
1084 Carmichael, H.-P. Chiu, K. D. Azarm, G. D. Haas, F. Amanat, J. Klingler, I. Baine, S.

1085 Arinsburg, J. C. Bandres, M. N. A. Siddiquey, R. M. Schilke, M. D. Woolard, H. Zhang,
1086 COVIDAR Argentina Consortium, A. J. Duty, T. A. Kraus, T. M. Moran, D. Tortorella, J.
1087 K. Lim, A. V. Gamarnik, C. E. Hioe, S. Zolla-Pazner, S. S. Ivanov, J. P. Kamil, F.
1088 Krammer, B. Lee, Quantifying Absolute Neutralization Titers against SARS-CoV-2 by a
1089 Standardized Virus Neutralization Assay Allows for Cross-Cohort Comparisons of COVID-
1090 19 Sera. *MBio* **12** (2021).

1091 66. S. Weiss, J. Klingler, C. Hioe, F. Amanat, I. Baine, S. Arinsburg, E. M. Kojic, J. Stoever, S.
1092 T. H. Liu, D. Jurczyszak, M. Bermudez-Gonzalez, V. Simon, F. Krammer, S. Zolla-Pazner,
1093 A High-Throughput Assay for Circulating Antibodies Directed Against the S Protein of
1094 Severe Acute Respiratory Syndrome Coronavirus 2. *J. Infect. Dis.* **222**, 1629–1634 (2020).

1095 67. James Bergstra, Daniel Yamins, and David Cox, "Making a Science of Model Search:
1096 Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures" in
1097 *Proceedings of the 30th International Conference on International Conference on Machine*
1098 *Learning* (PMLR, 2013)*28(1)*, pp. 115–123.

1099 68. B. Reynisson, B. Alvarez, S. Paul, B. Peters, M. Nielsen, NetMHCpan-4.1 and
1100 NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif
1101 deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* **48**,
1102 W449–W454 (2020).

1103 69. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7:
1104 improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

1105 70. R. Das, D. Baker, Macromolecular modeling with rosetta. *Annu. Rev. Biochem.* **77**, 363–
1106 382 (2008).

1107 71. H. Park, P. Bradley, P. Greisen Jr, Y. Liu, V. K. Mulligan, D. E. Kim, D. Baker, F. DiMaio,
1108 Simultaneous Optimization of Biomolecular Energy Functions on Features from Small
1109 Molecules and Macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).

1110 72. T. Kortemme, D. E. Kim, D. Baker, Computational alanine scanning of protein-protein
1111 interfaces. *Sci. STKE* **2004**, l2 (2004).

1112

1113 **Acknowledgments:**

1114 We gratefully acknowledge all data contributors, i.e., the Authors and their Originating
1115 laboratories responsible for obtaining the specimens, and their Submitting laboratories for
1116 generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which a
1117 portion of this research is based.

1118 We thank Vincenza Itri and Susan Zolla-Pazner at the Department of Medicine, the Division of
1119 Hematology and Medical Oncology, Icahn School of Medicine at Mount Sinai, New York, NY,
1120 USA, for generously allowing use of their Luminex machine in their laboratory for multiplexed
1121 antibody binding assays and providing technical supervision of reagent and equipment use.

We thank the members of the Mount Sinai COVID-19 Biobank Team listed here for their contributions to patient sample collection and processing: Charuta Agashe, Priyal Agrawal, Alara Akyatan, Kasey Alesso-Carra, Eziwoma Alibo, Kelvin Alvarez, Angelo Amabile, Carmen Argmann, Kimberly Argueta, Steven Ascolillo, Rasheed Bailey, Craig Batchelor, Noam D. Beckmann, Aviva G. Beckmann, Priya Begani, Jessica Le Berichel, Dusan Bogunovic, Swaroop Bose, Cansu Cimen Bozkus, Paloma Bravo, Mark Buckup, Larissa Burka, Sharlene Calorossi, Lena Cambron, Guillermo Carbonell, Gina Carrara, Mario A. Cedillo, Christie Chang, Serena Chang, Alexander W. Charney, Steven T. Chen, Esther Cheng, Jonathan Chien, Mashkura Chowdhury, Jonathan Chung, Phillip H. Comella, Dana Cosgrove, Francesca Cossarini, Liam Cotter, Arpit Dave, Travis Dawson, Bheesham Dayal, Diane Marie Del Valle, Maxime Dhainaut, Rebecca Dornfeld, Katie Dul, Melody Eaton, Nissan Eber, Cordelia Elaiho, Ethan Ellis, Frank Fabris, Jeremiah Faith, Dominique Falci, Susie Feng, Brian Fennessy, Marie Fernandes, Nataly Fishman, Nancy J. Francoeur, Sandeep Gangadharan, Daniel Geanon, Bruce D. Gelb, Benjamin S. Glicksberg, Sacha Gnjatic, Joanna Grabowska, Gavin Gyimesi, Maha Hamdani, Diana Handler, Jocelyn Harris, Matthew Hartnett, Sandra Hatem, Manon Herbinet, Elva Herrera, Arielle Hochman, Gabriel E. Hoffman, Jaime Hook, Laila Horta, Etienne Humblin, Suraj Jaladanki, Hajra Jamal, Jessica S. Johnson, Gurpawan Kang, Neha Karekar, Subha Karim, Geoffrey Kelly, Jong Kim, Seunghee Kim-Schulze, Edgar Kozlova, Arvind Kumar, Jose Lacunza, Alona Lansky, Dannielle Lebovitch, Brian Lee, Grace Lee, Gyu Ho Lee, Jacky Lee, John Leech, Lauren Lepow, Michael B. Leventhal, Lora E. Liharska, Katherine Lindblad, Alexandra Livanos, Bojan Losic, Rosalie Machado, Kent Madrid, Zafar Mahmood, Kelcey Mar, Thomas U. Marron, Glenn Martin, Robert Marvin, Shrisha Maskey, Paul Matthews, Katherine Meckel, Saurabh Mehandru, Miriam Merad, Cynthia Mercedes, Elyze Merzier, Dara Meyer, Gurkan Mollaoglu, Sarah Morris, Konstantinos Mouskas, Emily Moya, Naa-akomaah Yeboah, Girish Nadkarni, Kai Nie, Marjorie Nisenholtz, George Ofori-Amanfo, Kenan Onel, Merouane Ounadjela, Manishkumar Patel, Vishwendra Patel, Cassandra Pruitt, Adeeb Rahman, Shivani Rathi, Jamie Redes, Ivan Reyes-Torres, Alcina Rodrigues, Alfonso Rodriguez, Vladimir Roudko, Panagiotis Roussos, Evelyn Ruiz, Pearl Scalzo, Eric E. Schadt, Ieisha Scott, Robert Sebra, Hardik Shah, Mark Shervey, Pedro Silva, Nicole W. Simons, Melissa Smith, Alessandra Soares Schanoski, Juan Soto, Shwetha Hara Sridhar, Stacey-Ann Brown, Hiyab Stefanos, Meghan Straw, Robert Sweeney, Alexandra Tabachnikova, Collin Teague, Ryan Thompson, Manying Tin, Kevin Tuballes, Scott R. Tyler, Bhaskar Upadhyaya, Akhil Vaid, Verena Van Der Heide, Natalie Vaninov, Konstantinos Vlachos, Daniel Wacker, Laura Walker, Hadley Walsh, Wenhui Wang, Bo Wang, C. Matthias Wilk, Lillian Wilkins, Karen M. Wilson, Jessica Wilson, Hui Xie, Li Xue, Nancy Yi, Ying-chih Wang, Mahlet Yishak, Sabina Young, Alex Yu, Nina Zaks, Renyuan Zha.

Funding:

Pershing Square Foundation's COVID-19 RFP grant GCO# 20-1838-00001-01 (NB)

Icahn School of Medicine at Mount Sinai, Division of Hematology and Medical Oncology, seed funds (NB)

NIH/NCI Cancer Center Support Grant P30 CA008748 (BDG)

National Institute of Health (NIH) Public Health Service Institutional Research Training award AI07647 (MB)

Author contributions:

Conceptualization: CCB, MB, VR, BDG, ML, NB

Formal Analysis: CCB, MB, MT, EW, TO

Funding Acquisition: BDG, NB

Investigation: CCB, MB, LV, MT, EW, TO, DG, YB, DR, PK, JK, VR, DH

Methodology: CCB, MB, LV, MT, EW, TO, NV, ML, NB

Resources: KYO, KDS, GK, HA, NK, CM, RG, KN, DDV, DD, DR, JS, PSG, AC, MM, SKS, BL, AW, VS

Supervision: CCB, DC, NV, ML, NB

Visualization: CCB, MB, MT, EW, TO

Writing – Original Draft: CCB, MB, MT, EW

Writing – Review & Editing: CCB, MB, MT, EW, TO, YB, DH, KDS, GK, HA, NK, RG, DD, EC, PSG, BL, VS, DC, ML, NB

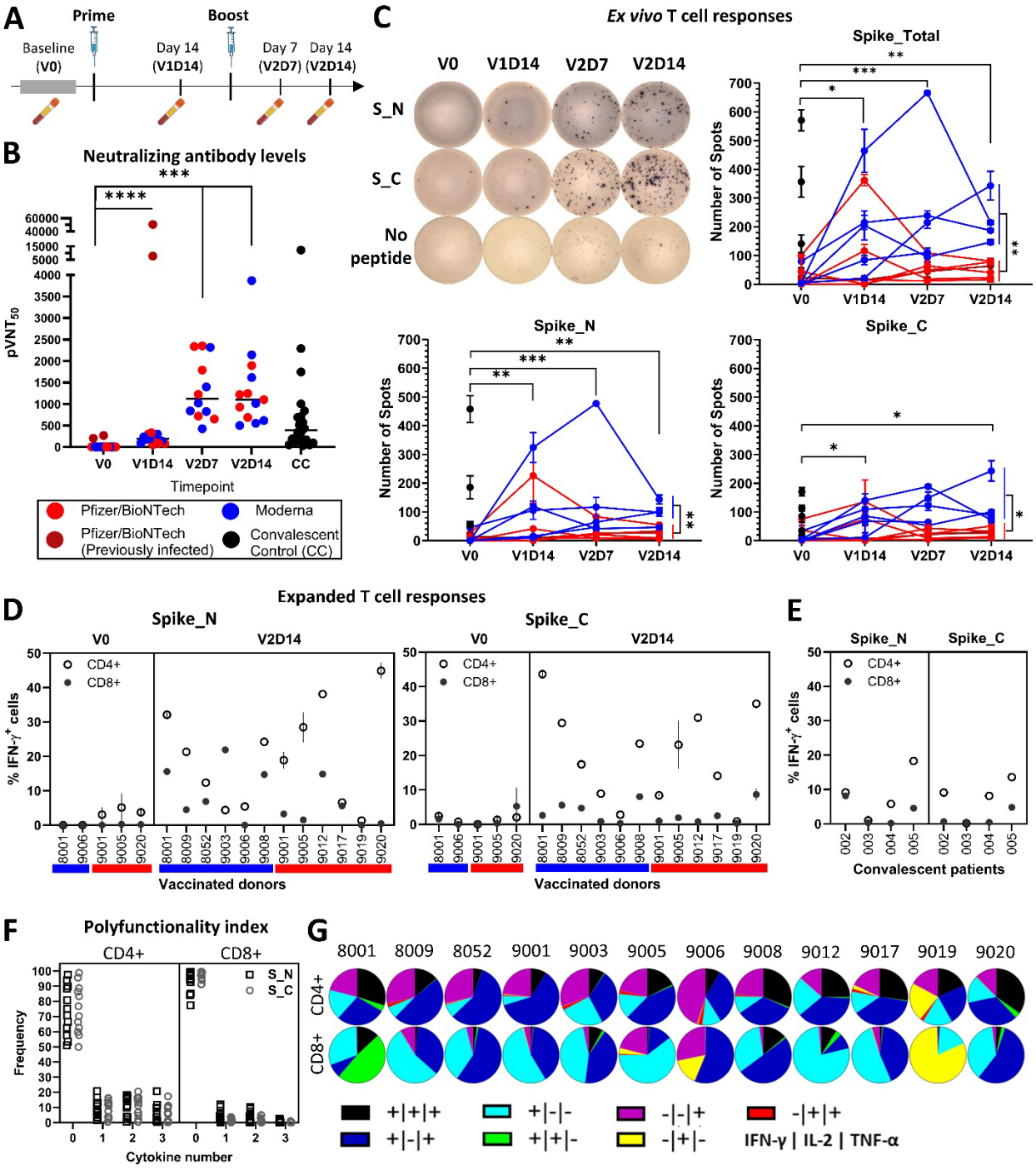
Competing interests:

CCB is a Bridge Fellow of the Parker Institute of Cancer Immunotherapy (PICI) and received research support. MB is a PICI Scholar. TO is an employee of Imprint Labs and a consultant for CDI Labs, Shennon Biotechnologies, and PopVax. BDG has received honoraria for speaking engagements from Merck, Bristol Meyers Squibb, and Chugai Pharmaceuticals; has received research funding from Bristol Meyers Squibb, Merck, and ROME Therapeutics; and has been a compensated consultant for Darwin Health, Merck, PMV Pharma, Shennon Biotechnologies, and Rome Therapeutics of which he is a co-founder. NB serves as an advisor/board member for Apricity, Break Bio, Carisma Therapeutics, CureVac, Genotwin, Novartis, Primevax, Rome Therapeutics, and Tempest Therapeutics; as a consultant for Genentech, Novartis, and ATP; receives research support from Dragonfly Therapeutics, Harbour Biomed Sciences, Regeneron Pharmaceuticals, and Ludwig Institute for Cancer Research; is an extramural member of PICI and receives research support. The remaining authors did not declare competing interests.

Data and materials availability:

The findings of a portion of this study (Fig. 5A-D) are based on 8,897,424 individual genome sequences and associated metadata published in GISAID's EpiCoV database (*46*) up to the following dates 2022-08-21, 2023-02-26, and 2024-03-21 via EPI_SET_240918oy. All sequences in this dataset are compared relative to hCoV-19/Wuhan/WIV04/2019 (WIV04), the official reference sequence employed by GISAID (EPI_ISL_402124). To view the contributors of each individual sequence with details such as accession number, Virus name, Collection date, Originating Lab and Submitting Lab and the list of Authors, visit 10.55876/gis8.240918oy. All other data needed to evaluate the conclusions of the paper are available in the main text or the supplementary materials. Python scripts used for model fitting are available upon request.

1204 **Figures:**



1205

1206 **Fig. 1. SARS-CoV-2 mRNA vaccine-induced adaptive immunity. (A)** Peripheral blood
1207 samples were collected from individuals receiving COVID-19 mRNA vaccines, mRNA-1273
1208 (by Moderna) or BNT162b2 (by Pfizer/BioNTech) longitudinally: before vaccination (V0), 14
1209 days after 1st dose (V1D14), 7 days after 2nd dose (V2D7) and 14 days after 2nd dose (V2D14),
1210 and from convalescent COVID-19 patients, who were not vaccinated. **(B)** Serum neutralization

1211 capacity was assessed by a pseudotype particle (pp) infection system, VSVΔG-Rluc bearing the
1212 SARS-CoV-2 D614G spike glycoprotein targeting 293T cells stably expressing ACE2 and
1213 TMPRSS2. 4-point nonlinear regression curves were used to calculate 50% pseudovirus
1214 neutralization titers (pVNT50) for vaccine recipients (n=12) and convalescent patients (n=20).
1215 Horizontal lines denote median pVNT50 values. Peripheral blood mononuclear cells (PBMCs,
1216 $2x10^5$ cells/well) were stimulated with pools of overlapping peptides, 15mers with 5 amino acid
1217 offsets, spanning the N-terminal (Spike_N or S_N) or C-terminal half (Spike_C or S_C) of Spike
1218 protein for 24 h and IFN-γ secretion was measured by ELISPOT. **(C)** Representative ELISPOT
1219 wells from a vaccinated donor (top left) and summary of ELISPOT data (n=11 vaccinated
1220 donors, BNT162b2 recipients in shown red, mRNA-1273 recipients in blue, and n=5
1221 convalescent patients, shown in black), where Spike_Total (top right) is the sum of spots
1222 acquired by Spike_N (bottom left) and Spike_C (bottom right) peptide pools. T cells were
1223 expanded following stimulation with Spike_N and Spike_C peptide pools. Antigen-specific
1224 cytokine production by expanded T cell subsets, CD4+ or CD8+, were measured by intracellular
1225 staining by flow cytometry. IFN-γ production by Spike-specific T cells **(D)** in vaccinated donors
1226 and **(E)** convalescent patients. **(F)** Polyfunctionality of Spike-specific T cells at V2D14 as
1227 demonstrated by % of T cells (y axis) co-expressing effector cytokines: IFN-γ, TNF-α, and IL-2
1228 (x axis). **(G)** Distribution of Spike-specific (Spike_Total) T cell responses in each vaccinated
1229 donor at V2D14. Spot numbers and cytokine+ cell frequencies were demonstrated after
1230 background subtraction. Statistical significance ($p < 0.05$) was evaluated by Wilcoxon matched-
1231 pairs test by comparing vaccination timepoints and Welch's t-test was used for comparing T cell
1232 responses elicited by mRNA-1273 vs BNT162b2.

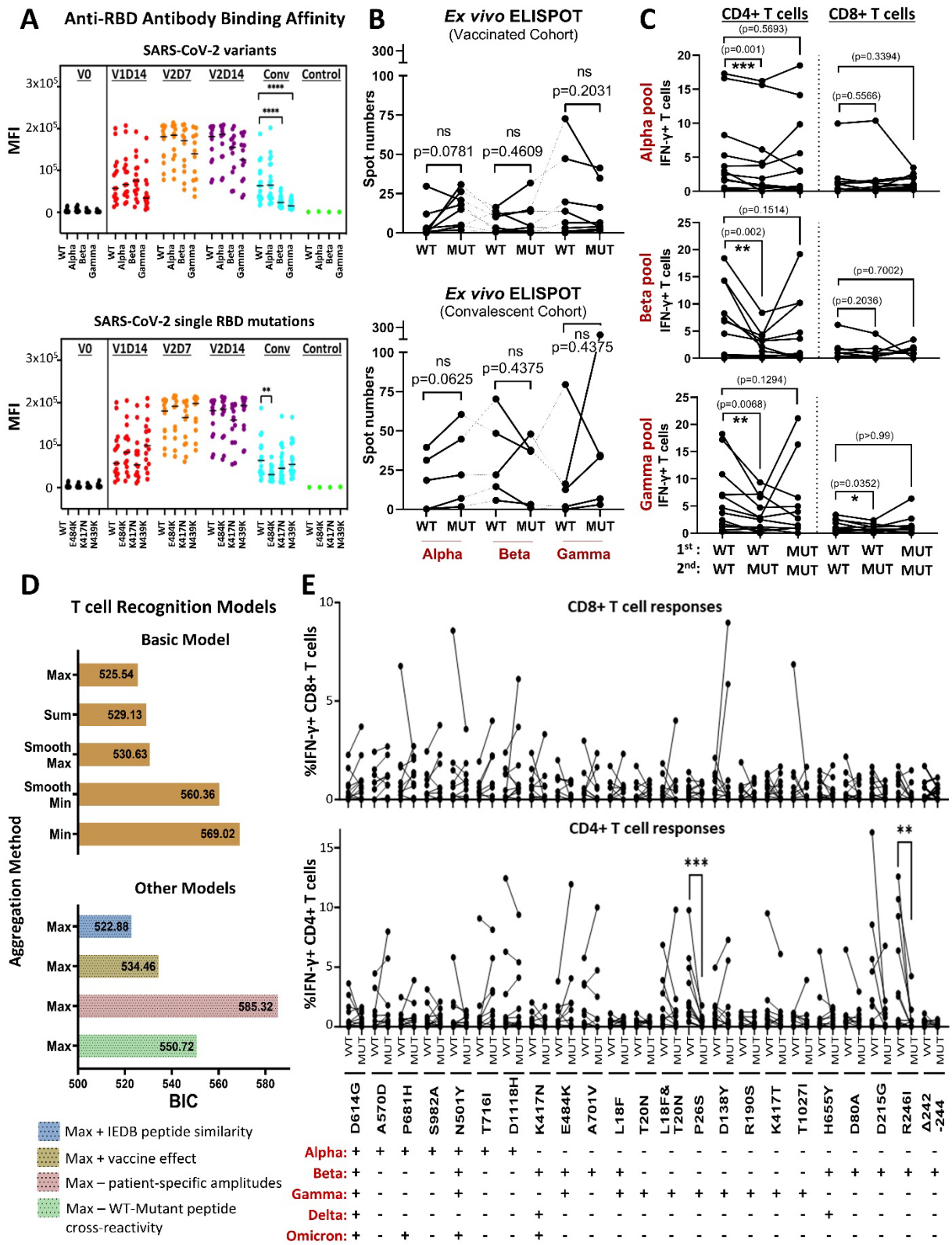


Fig. 2. Adaptive immune recognition of SARS-CoV-2 variants. (A) Antibody reactivity from vaccinated (before vaccination (V0), 14 days after 1st dose (V1D14), 7 days after 2nd dose

1236 (V2D7) and 14 days after 2nd dose (V2D14)), convalescent (conv), or pre-COVID-19 (control)
1237 serum to RBD were assessed by Luminex antibody binding assay where MagPlex-C
1238 Microspheres Regions were conjugated to recombinant wild-type (WT, Wuhan-1) and mutant
1239 RBD constructs (Alpha [N501Y], Beta [N501Y/K417N/E484K], and Gamma
1240 [N501Y/K417T/E484K] with mean fluorescence intensity (MFI) used as a readout for binding
1241 affinity. **(B)** Summary of ELISPOT data: peripheral blood mononuclear cells (PBMCs, $2x10^5$
1242 cells/well) from vaccinated donors (V2D14) or from convalescent patients were stimulated with
1243 pooled peptides covering the mutations found in Alpha, Beta and Gamma variants (listed in (E))
1244 or the corresponding WT sequences for 24 h and IFN-γ secretion was measured by ELISPOT.
1245 **(C)** V2D14 T cells from vaccinated donors were stimulated with variant or WT peptide pools
1246 and expanded prior to being re-stimulated with either the initial stimulation peptide pool
1247 (WT→WT, Mut→Mut) or the variant pool to measure cross-reactivity (WT→Mut). Antigen-
1248 specific cytokine production by expanded T cell subsets, CD4+ or CD8+, was measured by
1249 intracellular staining by flow cytometry. **(D)** Bars show Bayesian Information Criterion (BIC)
1250 values for different models of T cell reactivity shown in C, where lower BIC is better. The basic
1251 model includes the effect of peptide pool stimulations and WT-Mutant peptide cross-reactivity.
1252 Different peptide dominance models are shown on the y axis, which correspond to the 9mer
1253 aggregation function used. Performance of other models was also measured, blue: accounting for
1254 sequence similarity with IEDB epitopes, brown: including vaccination effect as the initial
1255 stimulation event, pink: excluding the effect of patient-specific amplitudes, green: excluding the
1256 effect of cross-reactivity between peptides from 1st and 2nd stimulation events. **(E)** Deconvolution
1257 of IFN-γ production by T cells in response to individual mutations within the peptide pools tested
1258 in C (WT→WT vs Mut→Mut). Statistical significance ($p < 0.05$) was evaluated by Wilcoxon
1259 matched-pairs test.

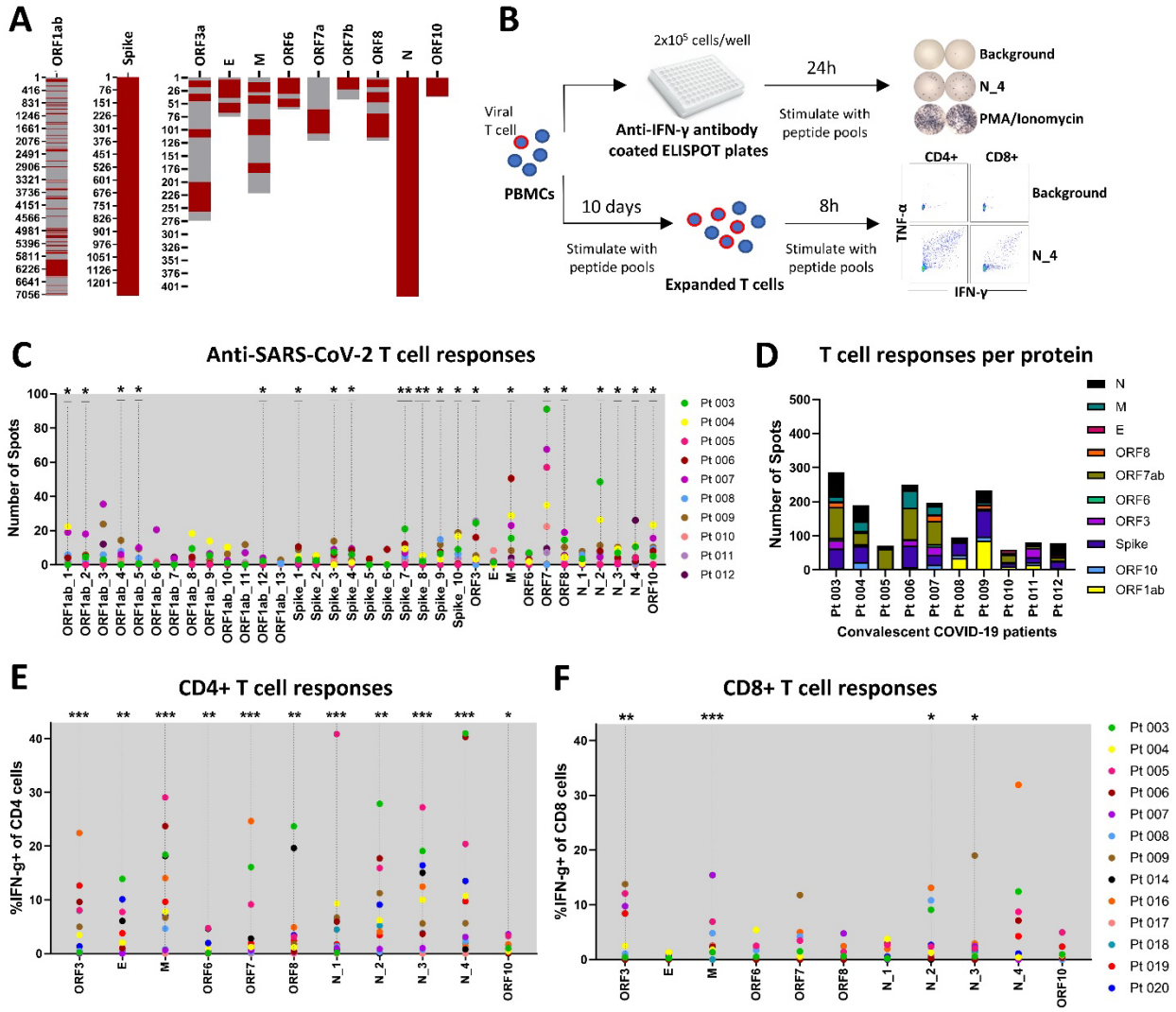


Fig. 3. T cell responses against non-spike SARS-CoV-2 proteins. (A) Schema demonstrating peptide selection for each protein. The y axis denotes amino acid residue number. Tested regions are shown in red. **(B)** Experimental design summary. Peripheral blood mononuclear cells (PBMCs, $2x10^5$ cells/well) from convalescent, unvaccinated patients (n=10) were stimulated for 24 h with pools of overlapping peptides, 15mers with 5 amino acid offsets, spanning each protein as indicated on the x axis. Pools contained no more than 25 peptides. IFN-γ secretion was measured by ELISPOT. Summary of ELISPOT data showing total T cell responses **(C)** per peptide pool and **(D)** per patient. T cells from the same convalescent, unvaccinated patient cohort (n=15) were expanded following stimulation with non-spike peptide pools. Antigen-specific cytokine production (IFN-γ+) by expanded T cell subsets, **(E)** CD4+ or **(F)** CD8+, was measured by intracellular staining by flow cytometry. Each dot corresponds to a patient. Normalized values were shown. Statistical significance ($p < 0.05$) was evaluated by the Wilcoxon matched-pairs test comparing DMSO vs peptide stimulation.

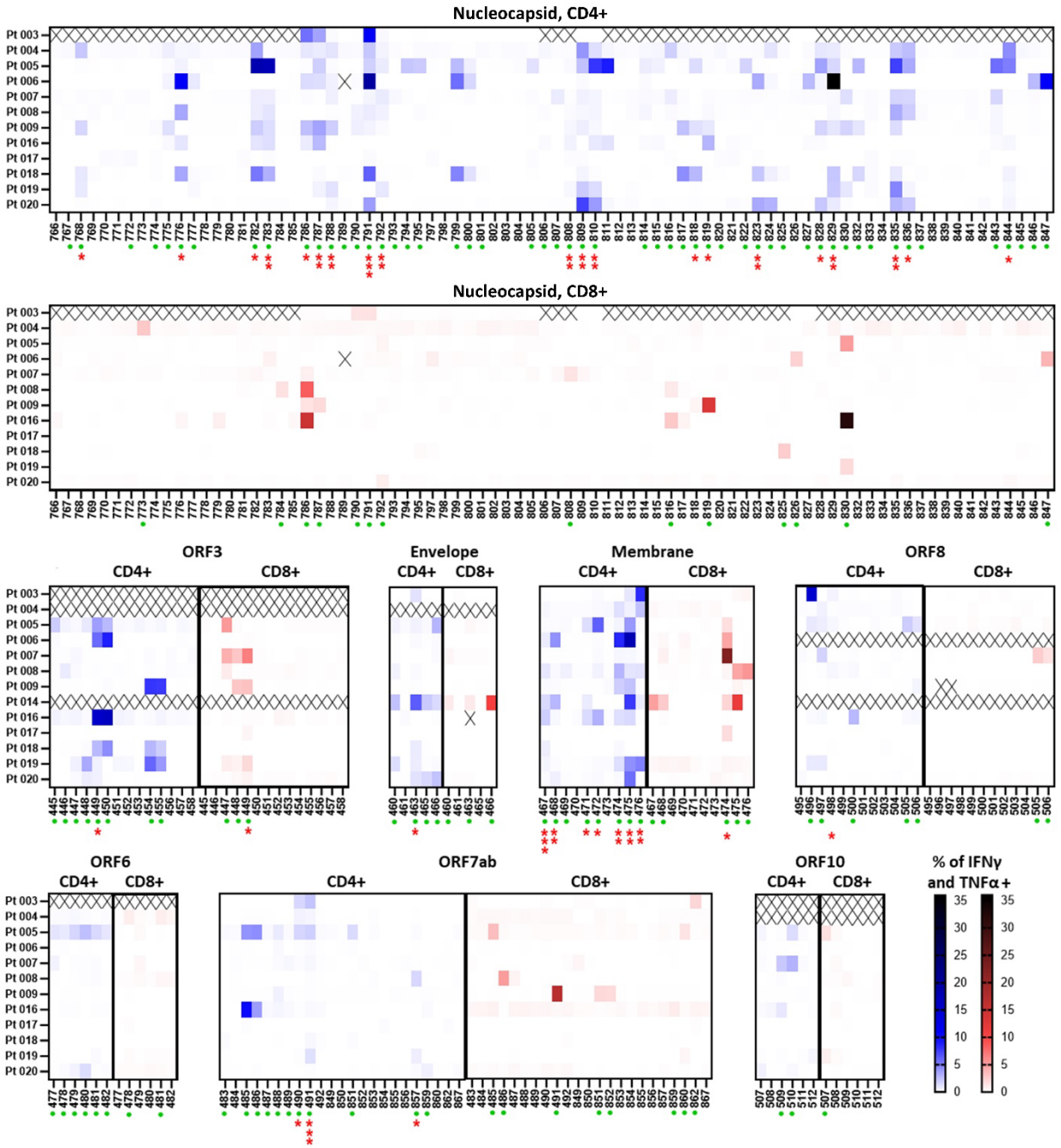


Fig. 4. Deconvolution of T cell responses against non-spike SARS-CoV-2 proteins. T cells from convalescent, unvaccinated patients were expanded following stimulation with non-spike peptide pools. Then, expanded T cells were re-stimulated by individual 15mers constituting the peptide pools. Antigen-specific cytokine production by expanded T cell subsets, CD4+ (in blue) or CD8+ (in red), was measured by intracellular staining by flow cytometry. Heat maps demonstrate the percentage of reactive, polyfunctional T cells (secreting both IFN-γ and TNF-α) after normalization (subtraction of background, DMSO stimulation). y and x axes indicate the patients and peptides tested, respectively. Peptide sequences are reported in table S2. X indicates that the data was not collected. Statistical significance ($p < 0.05$) for peptides inducing T cell

1285 responses across the tested population was evaluated by Wilcoxon matched-pairs test comparing
1286 DMSO vs peptide stimulation and denoted in red stars if significant (immunodominant peptides).
1287 A peptide was considered immunogenic, a “hit”, if in at least one patient, the % of reactive cells
1288 was greater than the paired DMSO % plus 3 times the standard deviation (>DMSO + 3SD) of all
1289 DMSO values across the population, denoted by green dots.
1290
1291

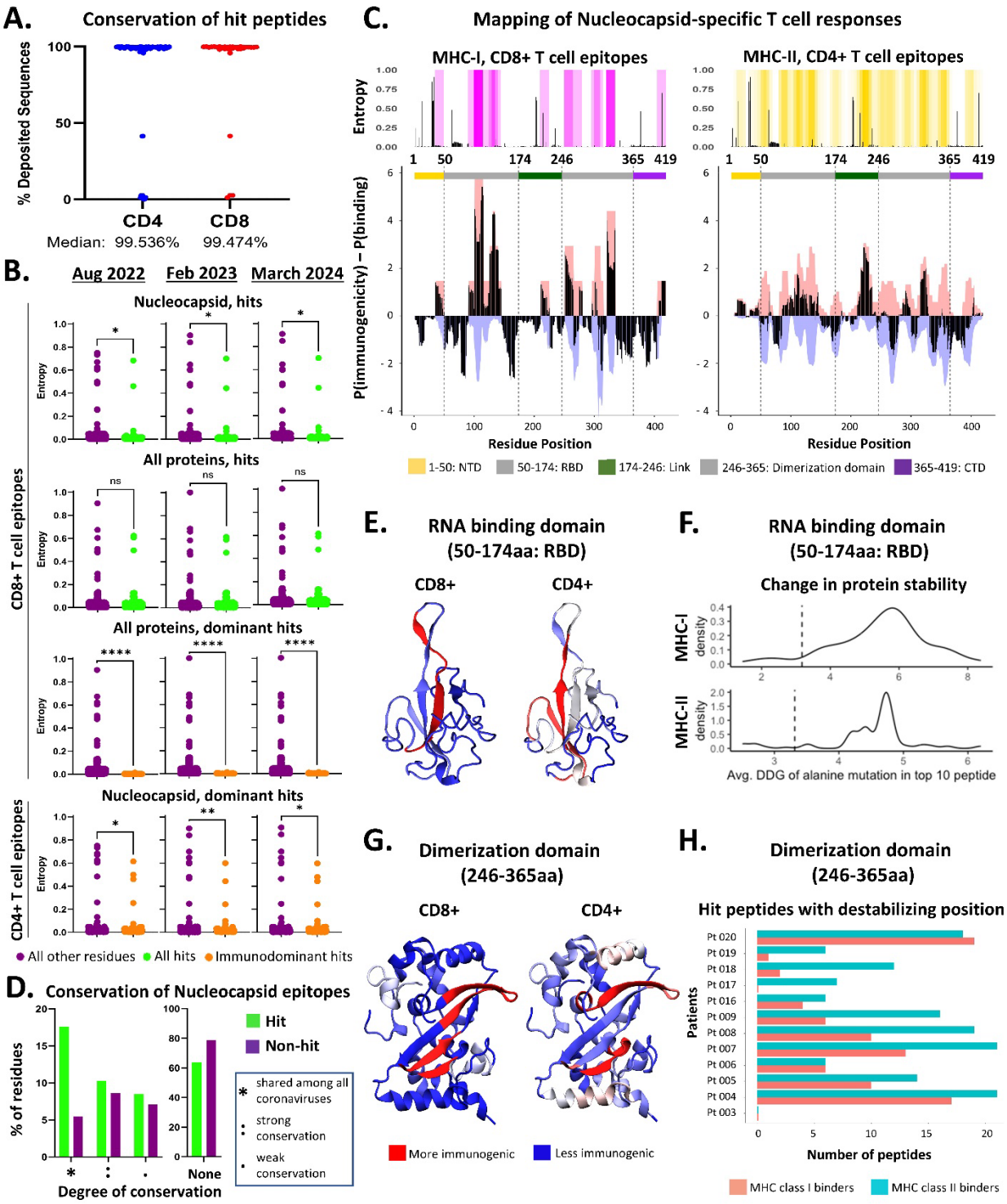
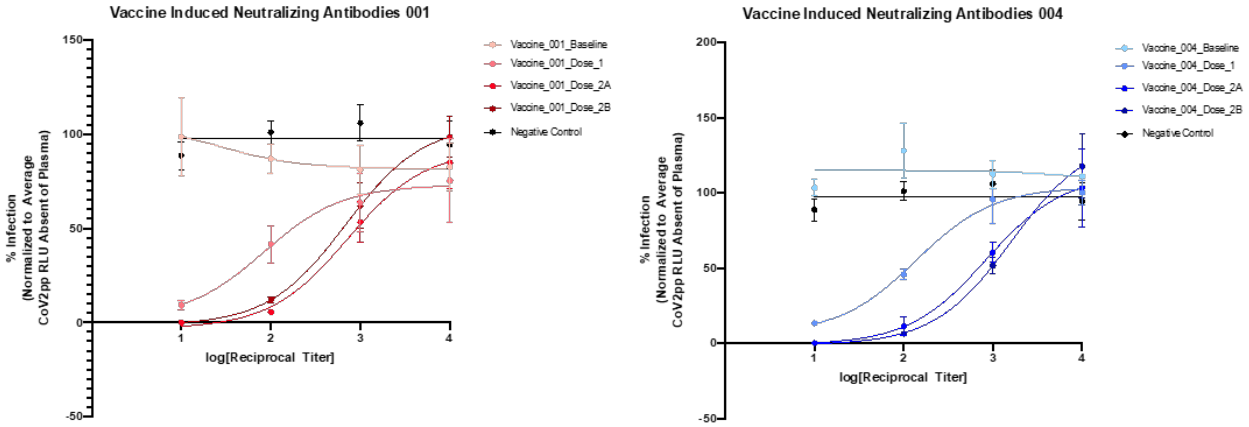


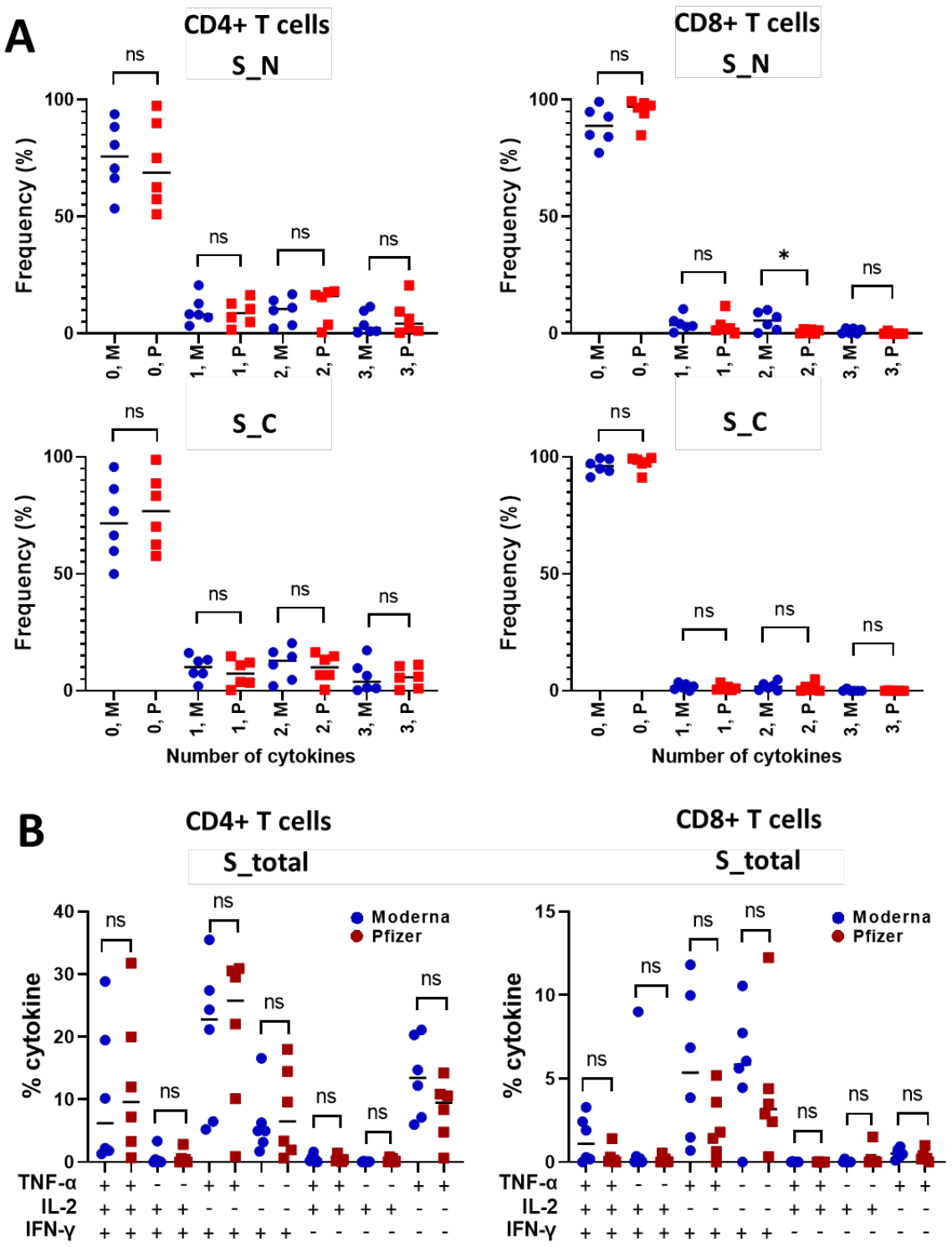
Fig. 5. Conservation of T cell epitopes. (A) Conservation of the "hit" T cell peptides, identified in fig. 4, was measured as percent of previously deposited GISAID sequences with an exact match to the reactive peptide. The median conservation percentages for the CD4 and CD8 "hit peptides" were 99.54% and 99.47%, respectively. **(B)** Entropy values for the residues found in

1298 “hit” peptides (green) or “hit” peptides that were significantly enriched across the test population
1299 (orange) were compared to the entropy values for all other residues (purple) in the ORFs tested
1300 in our study. Statistical significance was calculated by Welch's t-test. **(C)** Topography of
1301 immunogenic nucleocapsid residues is displayed. NTD: N-terminal domain, RBD: RNA binding
1302 domain, DD: dimerization domain, CTD: C-terminal domain. The black bars represent the
1303 difference between per residue immunogenicity (experimental, shaded in pink) and per residue
1304 binding value (predicted, shaded in lilac). Per residue immunogenicity values were generated by
1305 summing for each residue the normalized percentage of CD8 or CD4 T cells expressing both
1306 TNFα and IFNγ (>DMSO + 3SD) directed at each peptide for which the residue belongs. Per
1307 residue binding value corresponds to the frequency that a given residue appeared in a peptide
1308 predicted to bind to a patient MHC. The per residue values were scaled from 0-1 with 1
1309 representing the residue with the highest immunogenicity value or the most frequently included
1310 in a predicted binding peptide and 0 representing the residue with the lowest immunogenicity or
1311 least frequent. The per residue values for binding predictions were then multiplied by -1 to
1312 reverse the sign. The Y axis was then transformed to represent the log odds ratio of the
1313 probability of being immunogenic vs predicted binders by dividing by the background
1314 probability (1/number of residues in the nucleocapsid protein). The normalized entropy per
1315 amino acid (aa) codon was also aligned with nucleocapsid, black bars denoting the degree of
1316 entropy and heatmaps showcasing the intensity of sharing of immunogenic residues in our
1317 cohort. **(D)** Summary of conservation degree for nucleocapsid CD8+ hit sequences analyzed by
1318 CLUSTAL O (1.2.4) multiple sequence alignment for H-COV Nucleocapsid Protein: 229E
1319 (UniProt Accession: A0A127AU35), NL63 (UniProt Accession: Q06XQ2), HKU1 (UniProt
1320 Accession: Q5MQC6), OC43 (UniProt Accession: P33469), SARS_2 (UniProt Accession:
1321 P0DTC9). Hit peptides are marked in green and other tested sequences are in purple. Sharing
1322 was denoted as the following: “*” Residue shared among all coronaviruses in sequence
1323 alignment, “:” Conservation between groups of strongly similar properties, “.” Conservation
1324 between groups of weakly similar properties. **(E)** Mapping of immunogenic regions within RNA
1325 binding domain (RBD) on the 6YVO crystal structure. **(F)** The average change in protein
1326 stability (DDG) of RBD upon mutating each residue to alanine for the top ten most immunogenic
1327 peptides. Positive values indicate destabilizing mutations. The dashed line indicates the sliding
1328 window average (length 9 for MHC-I and 15 for MHC-II) across the nucleocapsid RBD. **(G)**
1329 Mapping of immunogenic regions within the dimerization domain onto the 6ZWO crystal
1330 structure. **(H)** Number of immunogenic peptides that contained at least one residue that could
1331 destabilize the interface when mutated to the alanine position (DDG>0 when mutated) was
1332 calculated for each patient.

Supplementary Figures



Supplementary Figure 1. Titration curves from two patients (001 and 004) for neutralizing antibody detection as summarized in Figure 1A.



Supplementary Figure 2. Comparison of effector cytokine secretion profile of anti-Spike T cells elicited by vaccination with mRNA-1273 (Moderna, in blue) or BNT162b2 (Pfizer/BioNTech, in red). Circulating anti-S T cells from V2D14 samples were expanded as described in Fig. 1D and cytokine formation was detected by intracellular flow cytometry. Significance was calculated by *t* test.

A

Mutation	Beta (B.1.351) South Africa	Gamma (P.1) Brazil	Alpha (B.1.1.7) UK	Delta (B.1.617.2 lineage) -	(B.1.1.529 and BA lineages) -	Mutated Sequence	WT Sequence
D614G	Y	Y	Y	Y	Y	PGTNTSNQVAVLYQGVNCTEVPVAIHADQ	PGTNTSNQVAVLYQDVNCTEVPVAIHADQ
A570D	N	N	Y	N	N	NKKFLPFQQFGRDIDDTTDAVRDPQTLEI	NKKFLPFQQFGRDIADTTDAVRDPQTLEI
P681H	N	N	Y	N	Y	GAGICASYQTQTNSHRRARSVASQSIIAY	GAGICASYQTQTNSPRRARSVASQSIIAY
S982A	N	N	Y	N	N	SNFGAISSVLNDILARLDKVEAEVQIDRL	SNFGAISSVLNDILSRLDKVEAEVQIDRL
N501Y	Y	Y	Y	N	Y	NCYFPLQSYGFQPTYGVGYQPYRVVVLSF	NCYFPLQSYGFQPTNGVGYQPYRVVVLSF
T716I	N	N	Y	N	N	ENSVAYSNNSIAIPINFTISVTTEILPVS	ENSVAYSNNSIAIPTNFTISVTTEILPVS
D1118H	N	N	Y	N	N	VTQRNFYEPQIITTHNTFVSGNCDVVIGI	VTQRNFYEPQIITTDNTFVSGNCDVVIGI
K417N	Y	N	N	Y	Y	RGDEVRQIAPGQTGNIADYNYKLPDDFTG	RGDEVRQIAPGQTGKIADYNYKLPDDFTG
E484K	Y	Y	N	N	N	TEIYQAGSTPCNGVKGFNCYFPLQSYGFQ	TEIYQAGSTPCNGVEGFNCYFPLQSYGFQ
A701V	Y	N	N	N	N	VASQSIIAYTMSLGVENSVAYSNNSIAIP	VASQSIIAYTMSLGAENSVAYSNNSIAIP
L18F	Y	Y	N	N	N	FLVLLPLVSSQCVNFTTRTQLPPAYTNSF	FLVLLPLVSSQCVNLTTRTQLPPAYTNSF
T20N	N	Y	N	N	N	VLLPLVSSQCVNLTNRTQLPPAYTNSFTR	VLLPLVSSQCVNLTTRTQLPPAYTNSFTR
P26S	N	Y	N	N	N	SSQCVNLTTRTQLPSAYTNSFTRGVYYPD	SSQCVNLTTRTQLPPAYTNSFTRGVYYPD
D138Y	N	Y	N	N	N	TNVVIKVCEFQFCNYPFLGVYYHKNNKSW	TNVVIKVCEFQFCNDPFLGVYYHKNNKSW
R190S	N	Y	N	N	N	LMDLEGKQGNFKNLSEFVFKNIDGYFKIY	LMDLEGKQGNFKNLREFVFKNIDGYFKIY
K417T	N	Y	N	N	N	RGDEVRQIAPGQTGTIADYNYKLPDDFTG	RGDEVRQIAPGQTGKIADYNYKLPDDFTG
T1027I	N	Y	N	N	N	IRAAEIRASANLAAIKMSECVLGQSKRVD	IRAAEIRASANLAATKMSECVLGQSKRVD
H655Y	N	Y	N	N	Y	NVFQTRAGCLIGAEYVNNSYECDIPIGAG	NVFQTRAGCLIGAEHVNNSYECDIPIGAG
D80A	Y	N	N	N	N	HAIHVSGTNGTKRFANPVLPFNDGVYFAS	HAIHVSGTNGTKRFDNPVLPFNDGVYFAS
D215G	Y	N	N	N	N	FKIYSKHTPINLVRGLPQGFSALEPLVDL	FKIYSKHTPINLVRDLPQGFSALEPLVDL
R246I	Y	N	N	N	N	GINITRFQTLLALHISYLTPGDSSSGWTA	GINITRFQTLLALHRSYLTPGDSSSGWTA
Δ242-244	Y	N	N	N	N	IGINITRFQTLHRSYLTPGDSSSGW	IGINITRFQTLLALHRSYLTPGDSSSGW
T20N&L18F	N/Y	Y	N	N	N	VLLPLVSSQCVNFTNRTQLPPAYTNSFTR	VLLPLVSSQCVNLTTRTQLPPAYTNSFTR

B

Mutation	Frequency	Presence
S:N501Y		UK, SA, Brazil

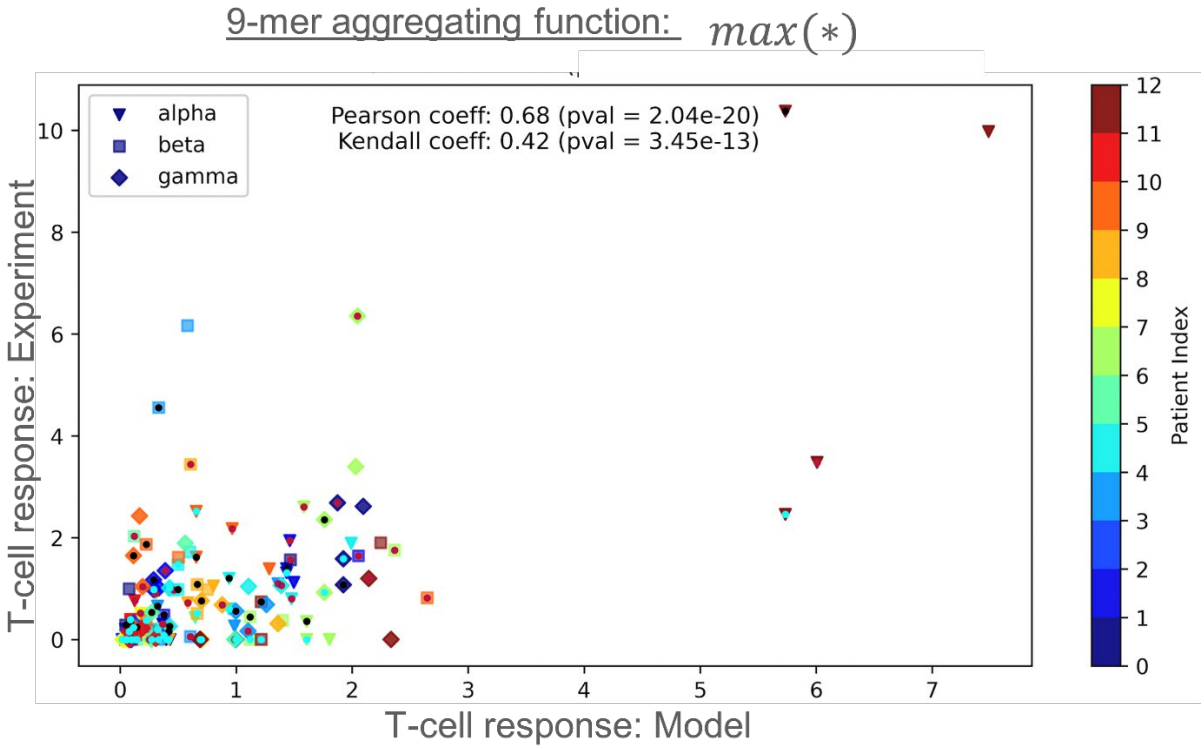
```
 490       500       510
NCYFPLQSYGFQPTNGVGYQPYRVVVLSF

NCYFPLQSYGFQPTNGVGYQPYRVVVLSF
Spike-99)    PLQSYGFQPTNGVGY
Spike-100)       GFQPTNGVGYQPYRV
Spike-101)           NGVGYQPYRVVVLSF
```

WT N501Y

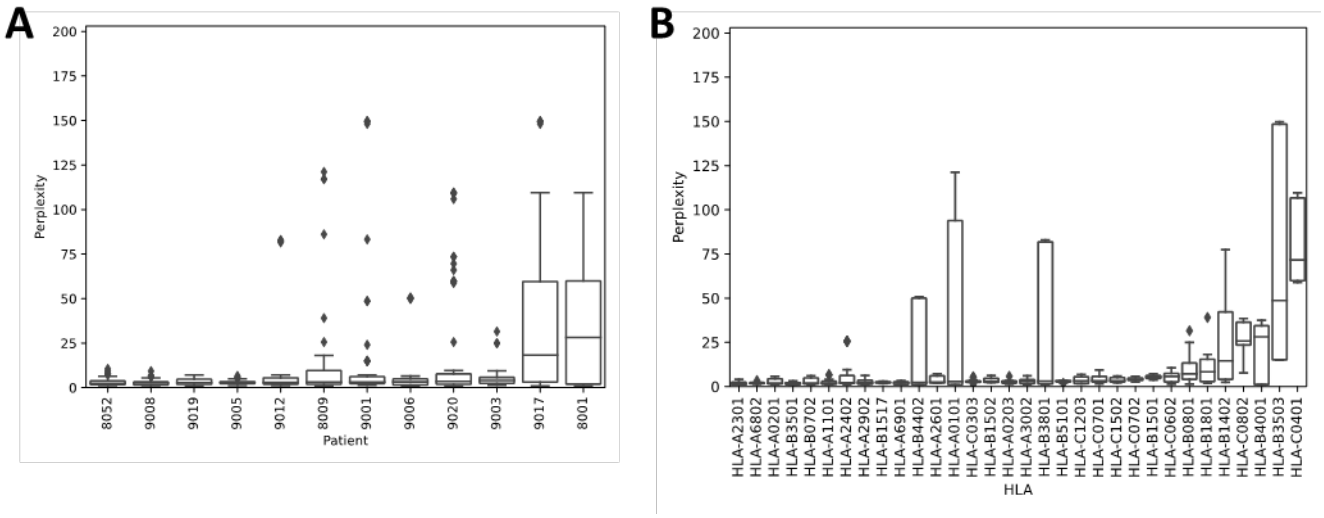
```
NCYFPLQSYGFQPTYGVGYQPYRVVVLSF
S:501Y-1)    PLQSYGFQPTYGVGY
S:501Y-2)       GFQPTYGVGYQPYRV
S:501Y-3)           YGVGYQPYRVVVLSF
```

Mutant N501Y

Supplementary Figure 3. A. Mutated and corresponding WT (Wuhan-1) sequences tested in figure 2. Red squares indicate the mutations included in clade pools indicated as Alpha, Beta, Gamma pools. **B.** Example peptide design for each mutation and its control sequence. A maximum of 14 upstream and downstream amino acid sequences surrounding the mutated amino acid were included and overlapping peptides were pooled per mutation and the corresponding WT sequence.

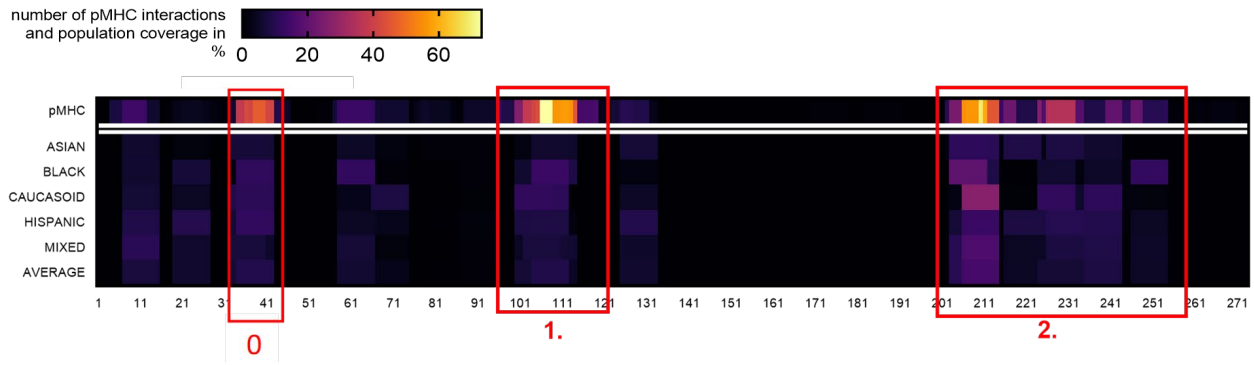


Supplementary Figure 4. The predictive capacity of modelling: modelling vs experimental data. Plot shows the result of optimizing the model using the maximum as the 9-mer aggregating function. Model responses are shown on the x-axis and experimental responses on the y-axis. Each mutation pool is denoted (top left) and each patient's data is color coded (right).



Supplementary Figure 5. Distribution of perplexity, the effective number of peptides that can be recognized among all the ninemers in a pool of peptides. **A.** For each patient we plot the distribution of perplexity of the ninemers, as computed over different HLA alleles of the patient, and the stimulation and restimulation peptide pools. The median perplexity over all conditions (patient, HLA, pool) is 3.07. **B.** Several HLA types are characterized by higher perplexity: HLA-C0802 (median P=25.77), HLA-B4001 (median P=28.14), HLA-B3503 (median P=48.64) and HLA-C0401 (median P=71.54)

>YP009724391_ORF3a

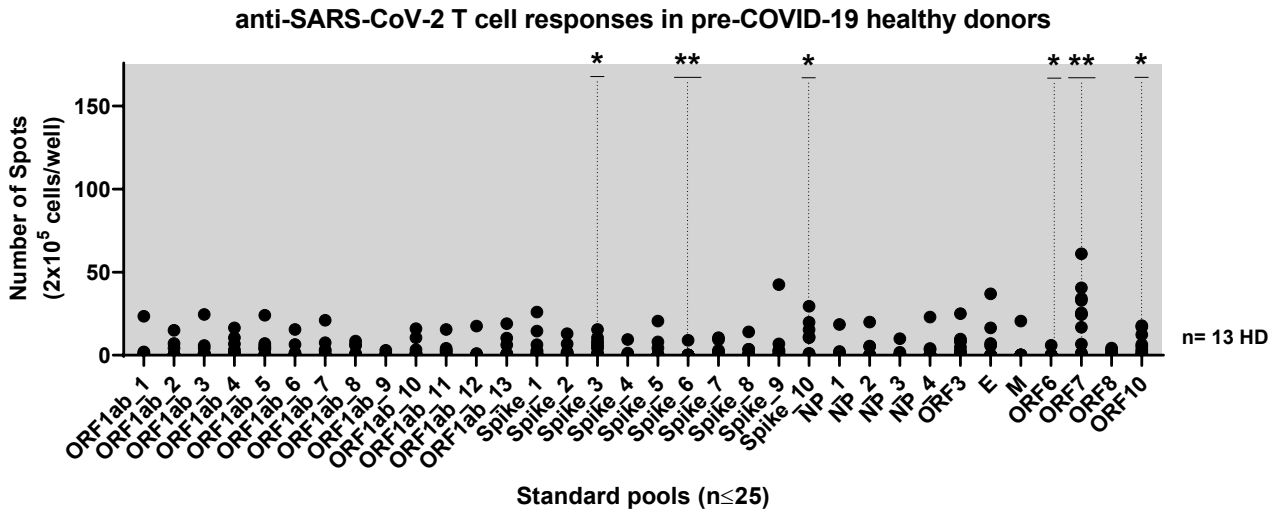



```
        10         20         30         40         50         60
MDLFMRIFTI GTVTLKQGEI KDATPSDFVR ATATIPIQAS LPFGWLIVGV ALLAVFQSAS

        70         80         90        100        110        120
KIITLKKRWQ LALSKGVHFV CNLLLLFVTV YSHLLLVAAG LEAPFLYLYA LVYFLQSINF

       130        140        150        160        170        180
VRIIMRLWLC WKCRSKNPLL YDANYFLCWH TNCYDYCIPY NSVTSSIVIT SGDGTTSPIS

       190        200        210        220        230        240
EHDYQIGGYT EKWESGVKDC VVLHSYFTSD YYQLYSTQLS TDTGVEHVTF FIYNKIVDEP

       250        260        270
EEHVQIHTID GSSGVVNPVM EPIYDEPTTT TSVPL
```

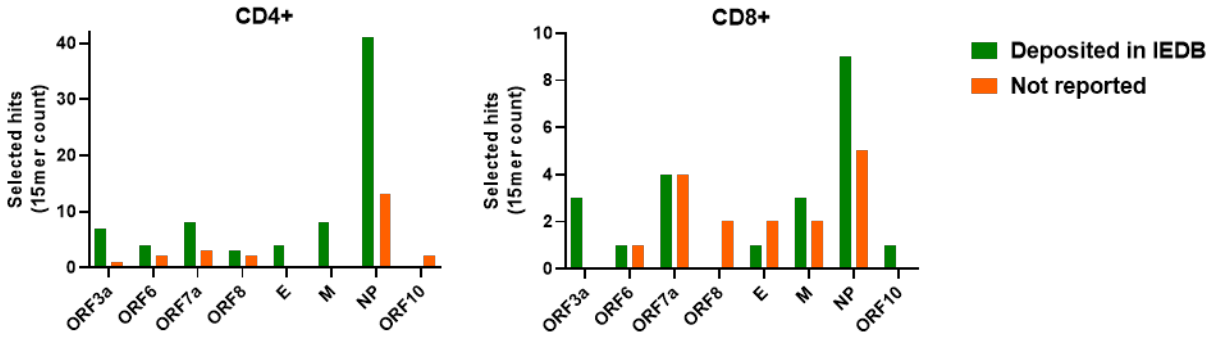
<u>OLP Design</u> (total # = 13)

```
      0.    TATIPIQASLPFGWL

ORF3a-0     TATIPIQASLPFGWL

      1.    GLEAPFLYLYALVYF

ORF3a-1     GLEAPFLYLYALVYF

      2.    VLHSYFTSDYYQLYSTQLSTDTGVEHVTFFIYNKIVDEPEEHVQIHTIDGSSGVV

ORF3a-2.1   VLHSYFTSDYYQLYS
ORF3a-2.2       YFTSDYYQLYSTQLS
ORF3a-2.3           DYYQLYSTQLSTDTG
ORF3a-2.4               LYSTQLSTDTGVEHV
ORF3a-2.5                   QLSTDTGVEHVTFFI
ORF3a-2.6                       DTGVEHVTFFIYNKI
ORF3a-2.7                           EHVTFFIYNKIVDEP
ORF3a-2.8                               FFIYNKIVDEPEEHV
ORF3a-2.9                                   NKIVDEPEEHVQIHT
ORF3a-2.10                                      DEPEEHVQIHTIDGS
ORF3a-2.11                                          EHVQIHTIDGSSGVV
```

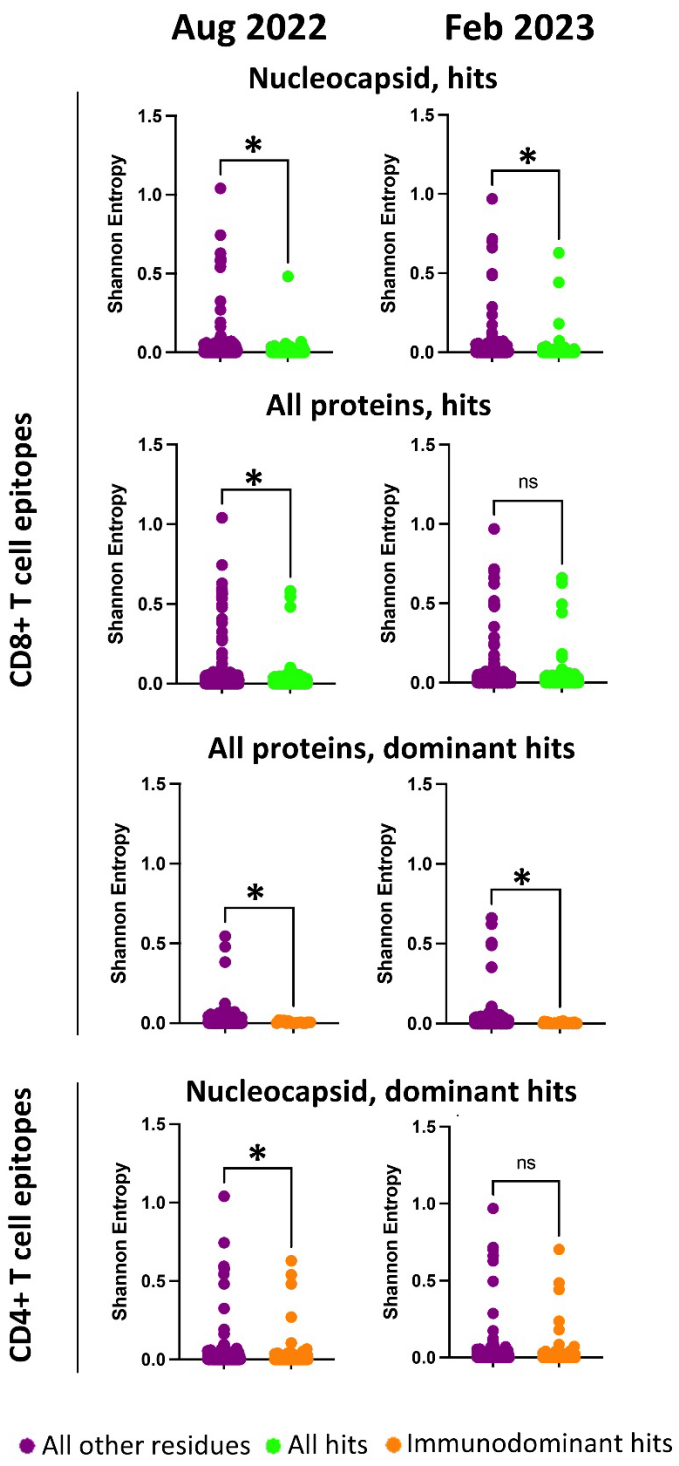
Supplementary Figure 6. ORF3a exemplifying peptide selection strategy. Regions with enhanced peptide-MHC interactions were selected to maximize population coverage, using IEDB's population coverage tool (http://tools.iedb.org/population/help/#by_ethnicity).



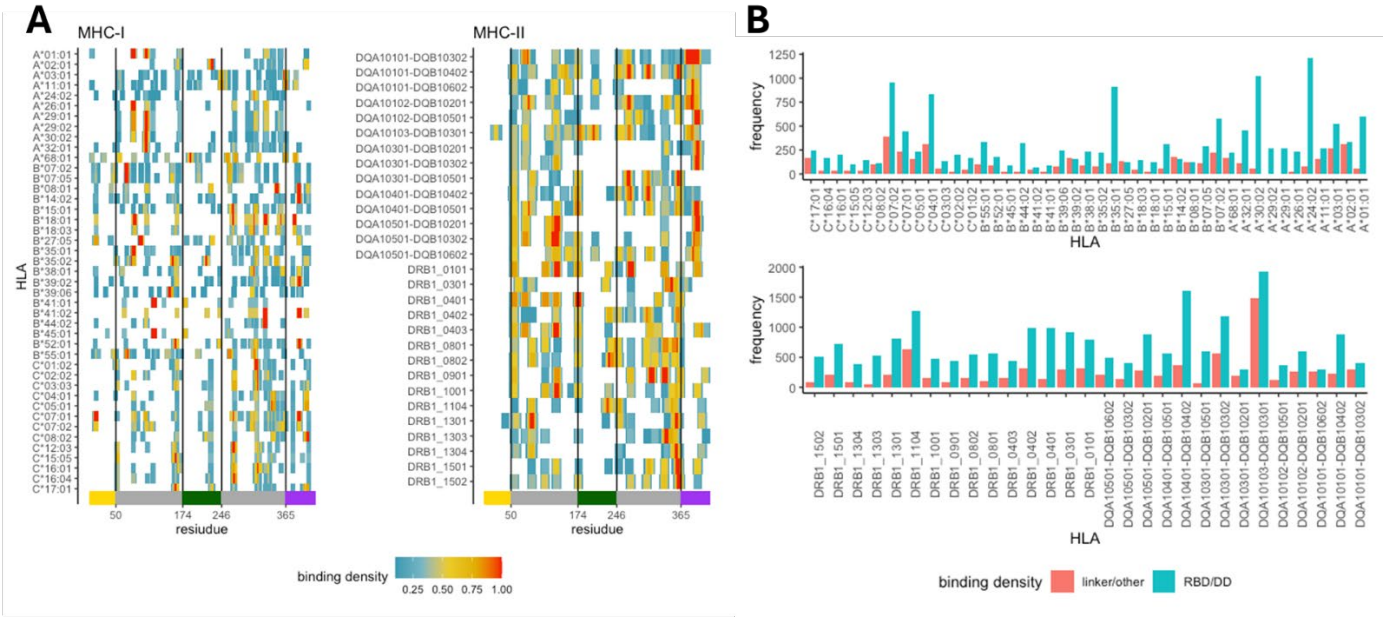
Supplementary Figure 7. SARS-CoV-2-specific T cell responses were measured *ex vivo* utilizing peripheral blood mononuclear cells (PBMCs, 2x105 cells/well) from COVID-19-naïve healthy donors (samples collected in 2019 and prior) were stimulated for 24 h with pools of overlapping peptides (max 25 peptides per pool as detailed in Figure 3: 15mers with 5 amino acids offset, spanning each protein as indicated on the x axis). IFN-γ secretion was measured by ELISPOT. Summary of ELISPOT data showing total T cell responses per peptide pool.



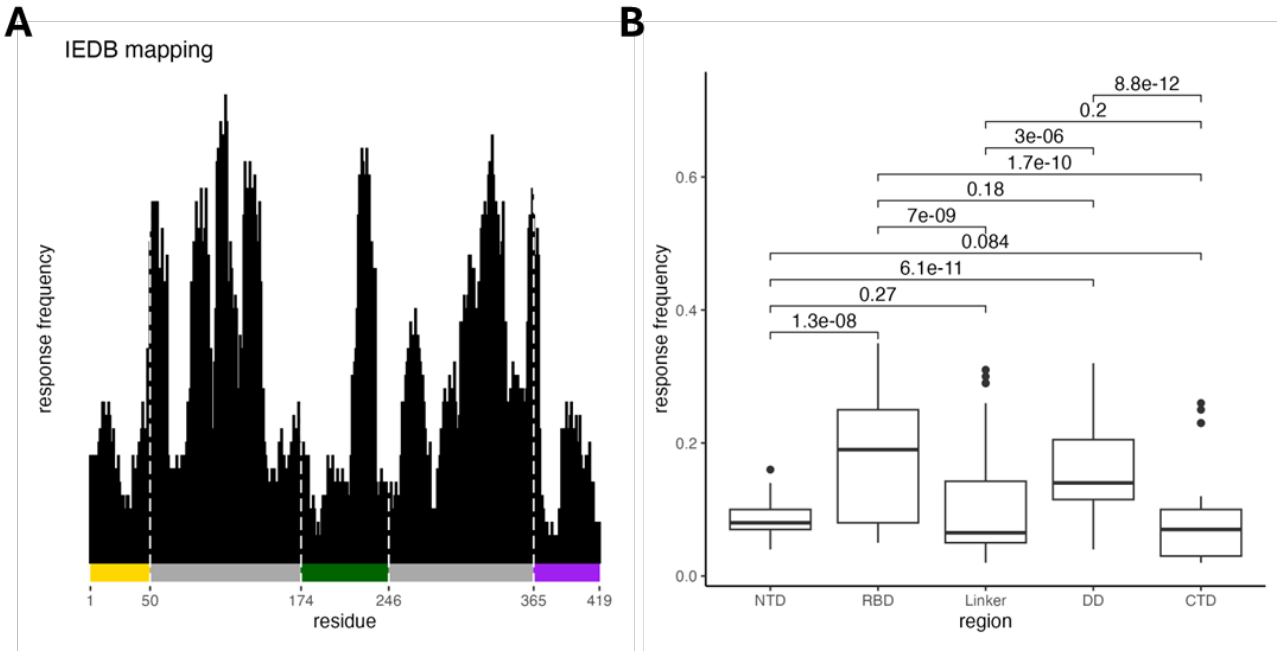
Supplementary Figure 8. The novelty of the immunogenic "hit" peptides eliciting CD4 and CD8 T cell responses were determined by predicting possible minimal epitopes using donor MHC alleles for each hit. A 15mer was considered novel if none of its predicted binders, across all patient MHC alleles, were present in IEDB. 23 out of 98 and 16 out of 38 for CD4 and CD8 response eliciting hit peptides (15mers), respectively, were deemed novel.



Supplementary Figure 9. The average Shannon entropy values were obtained from Nextstrain as of August 2022 or February 2023 and demonstrated for the residues found in “hit” peptides (green; denoted by green dot in fig. 4) or “hit” peptides that were significantly enriched across test population (orange; denoted by red stars in fig. 4) were compared to the average for all other residues (purple) in the ORFs tested in our study. Statistical significance was calculated by Welch's t-test.



Supplementary Figure 10. Number of predicted peptides by MHC-I and MHC-II alleles of convalescent patients tested in fig. 4. **A.** The number of appearances was normalized with respect to each allele with red indicating the most shared residues for peptides predicted for that allele and blue indicating the least common. White spaces signify that the given residue did not appear in a predicted peptide for that allele. **B.** The summary of targeted domains for each allele. Blue bars indicate the number of targeted residues derived from the highly structured regions, e.g. RNA binding domain (RBD) or the dimerization domain (DD). The pink bar indicates regions outside of those two.



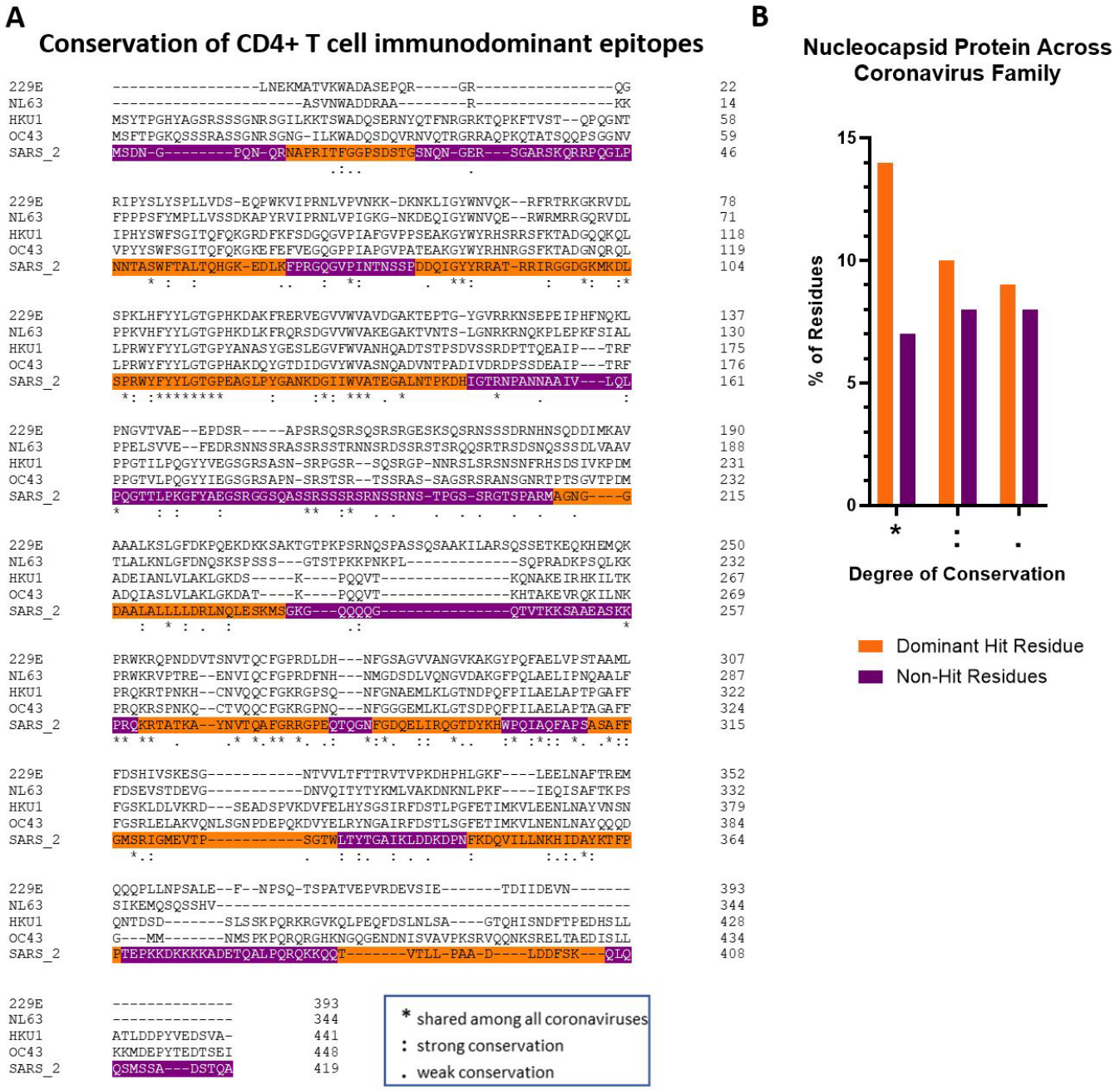
Supplementary Figure 11. Mapping of immunogenic Nucleocapsid peptides from IEDB. **A.** Residues indicate the following regions, yellow (1-50aa): N-Terminal Domain (NDT), grey (50-174aa): RNA binding domain (RBD), green (174-246aa): linker region, grey (246-365aa): dimerization domain (DD), purple (365-419aa): C-Terminal Domain (CTD). **B.** Quantification and frequency comparison of immunogenic peptides across different regions. The statistical significance of differences between means was tested using the Wilcoxon ranked sum test.

Conservation of CD8+ T cell hit epitopes

```
229E      -----------------LNEKMATVKWADASEPQR-----GR----------------QG     22
NL63      ----------------------ASVNWADDRAA--------R----------------KK     14
HKU1      MSYTPGHYAGSRSSSGNRSGILKKTSWADQSERNYQTFNRGRKTQPKFTVST--QPQGNT     58
OC43      MSFTPGKQSSSRASSGNRSGNG-ILKWADQSDQVRNVQTRGRRAQPKQTATSQQPSGGNV     59
SARS_2    MSDN-G--------PQN-QRNAPRITFGGPSDSTGSNQN-GER---SGARSKQRRPQGLP     46
                                 .:..           .

229E      RIPYSLYSPLLVDS-EQPWKVIPRNLVPVNKK-DKNKLIGYWNVQK--RFRTRKGKRVDL     78
NL63      FPPPSFYMPLLVSSDKAPYRVIPRNLVPIGKG-NKDEQIGYWNVQE--RWRMRRGQRVDL     71
HKU1      IPHYSWFSGITQFQKGRDFKFSDGQGVPIAFGVPPSEAKGYWYRHSRRSFKTADGQQKQL    118
OC43      VPYYSWFSGITQFQKGKEFEFVEGQGPPIAPGVPATEAKGYWYRHNRGSFKTADGNQRQL    119
SARS_2    NNTASWFTALTQHGK-EDLKFPRGQGVPINTNSSPDDQIGYYRRAT-RRIRGGDGKMKDL    104
              * :  :         ..   :  *:       .  **:        :   *:  :*

229E      SPKLHFYYLGTGPHKDAKFRERVEGVVWVAVDGAKTEPTG-YGVRRKNSEPEIPHFNQKL    137
NL63      PPKVHFYYLGTGPHKDLKFRQRSDGVVWVAKEGAKTVNTS-LGNRKRNQKPLEPKFSIAL    130
HKU1      LPRWYFYYLGTGPYANASYGESLEGVFWVANHQADTSTPSDVSSRDPTTQEAIP---TRF    175
OC43      LPRWYFYYLGTGPHAKDQYGTDIDGVYWVASNQADVNTPADIVDRDPSSDEAIP---TRF    176
SARS_2    SPRWYFYYLGTGPEAGLPYGANKDGIIWVATEGALNTPKDHIGTRNPANNAAIV---LQL    161
           *: :********     :    :*: *** . *          *    .        :

229E      PNGVTVAE--EPDSR-----APSRSQSRSQSRSRGESKSQSRNSSSDRNHNSQDDIMKAV    190
NL63      PPELSVVE--FEDRSNNSSRASSRSSTRNNSRDSSRSTSRQQSRTRSDSNQSSSDLVAAV    188
HKU1      PPGTILPQGYYVEGSGRSASN-SRPGSR--SQSRGP-NNRSLSRSNSNFRHSDSIVKPDM    231
OC43      PPGTVLPQGYYIEGSGRSAPN-SRSTSR--TSSRAS-SAGSRSRANSGNRTPTSGVTPDM    232
SARS_2    PQGTTLPKGFYAEGSRGGSQASSRSSSRSRNSSRNS-TPGS-SRGTSPARMAGNG----G    215
          *    : :    :         **  :*  . .    .  . .   .  .   .

229E      AAALKSLGFDKPQEKDKKSAKTGTPKPSRNQSPASSQSAAKILARSQSSETKEQKHEMQK    250
NL63      TLALKNLGFDNQSKSPSSS---GTSTPKKPNKPL-------------SQPRADKPSQLKK    232
HKU1      ADEIANLVLAKLGKDS-----K----PQQVT---------------KQNAKEIRHKILTK    267
OC43      ADQIASLVLAKLGKDAT----K----PQQVT---------------KHTAKEVRQKILNK    269
SARS_2    DAALALLLLDRLNQLESKMSGKG---QQQQG---------------QTVTKKSAAEASKK    257
             :  * : .  :            .:                              *

229E      PRWKRQPNDDVTSNVTQCFGPRDLDH---NFGSAGVVANGVKAKGYPQFAELVPSTAAML    307
NL63      PRWKRVPTRE--ENVIQCFGPRDFNH---NMGDSDLVQNGVDAKGFPQLAELIPNQAALF    287
HKU1      PRQKRTPNKH--CNVQQCFGKRGPSQ---NFGNAEMLKLGTNDPQFPILAELAPTPGAFF    322
OC43      PRQKRSPNKQ--CTVQQCFGKRGPNQ---NFGGGEMLKLGTSDPQFPILAELAPTAGAFF    324
SARS_2    PRQKRTATKA--YNVTQAFGRRGPEQTQGNFGDQELIRQGTDYKHWPQIAQFAPSASAFF    315
          ** **  .     .* *.** *. .:   *:*.  ::  *..   :* :*:: *. .*::

229E      FDSHIVSKESG-----------NTVVLTFTTRVTVPKDHPHLGKF----LEELNAFTREM    352
NL63      FDSEVSTDEVG-----------DNVQITYTYKMLVAKDNKNLPKF----IEQISAFTKPS    332
HKU1      FGSKLDLVKRD---SEADSPVKDVFELHYSGSIRFDSTLPGFETIMKVLEENLNAYVNSN    379
OC43      FGSRLELAKVQNLSGNPDEPQKDVYELRYNGAIRFDSTLSGFETIMKVLNENLNAYQQQD    384
SARS_2    GMSRIGMEVTP-----------SGTWLTYTGAIKLDDKDPNFKDQVILLNKHIDAYKTFP    364
            *.:                  .   : :.  : . .    :        :.:.*:

229E      QQQPLLNPSALE--F--NPSQ-TSPATVEPVRDEVSIE-------TDIIDEVN-------    393
NL63      SIKEMQSQSSHV-----------------------------------------------    344
HKU1      QNTDSD-------SLSSKPQRKRGVKQLPEQFDSLNLSA----GTQHISNDFTPEDHSLL    428
OC43      G---MM-------NMSPKPQRQRGHKNGQGENDNISVAVPKSRVQQNKSRELTAEDISLL    434
SARS_2    PTEPKKDKKKKADETQALPQRQKKQQT-------VTLL-PAA-D----LDDFSK---QLQ    408

229E      --------------          393
NL63      --------------          344
HKU1      ATLDDPYVEDSVA-          441
OC43      KKMDEPYTEDTSEI          448
SARS_2    QSMSSA---DSTQA          419
```

* shared among all coronaviruses
: strong conservation
. weak conservation

Supplementary Figure 12. CLUSTAL O(1.2.4) multiple sequence alignment analysis for H-COV Nucleocapsid Protein: 229E (UniProt Accession: A0A127AU35), NL63 (UniProt Accession: Q06XQ2), HKU1 (UniProt Accession: Q5MQC6), OC43 (UniProt Accession: P33469), SARS_2 (UniProt Accession: P0DTC9). Hit peptides inducing CD8+ T cell responses are marked by green and other tested sequences are in purple. Sharing was denoted as the following: "*" Residue shared among all coronaviruses in sequence alignment, ":" Conservation between groups of strongly similar properties, "." Conservation between groups of weakly similar properties.



Supplementary Figure 13. A. CLUSTAL O(1.2.4) multiple sequence alignment analysis for H-COV Nucleocapsid Protein: 229E (UniProt Accession: A0A127AU35), NL63 (UniProt Accession: Q06XQ2), HKU1 (UniProt Accession: Q5MQC6), OC43 (UniProt Accession: P33469), SARS_2 (UniProt Accession: P0DTC9). **B.** Summary of conservation degree. Immunodominant peptides inducing CD4+ T cell responses are marked by orange and other tested sequences are in purple. Sharing was denoted as the following: "*" Residue shared among all coronaviruses in sequence alignment, ":" Conservation between groups of strongly similar properties, "." Conservation between groups of weakly similar properties.