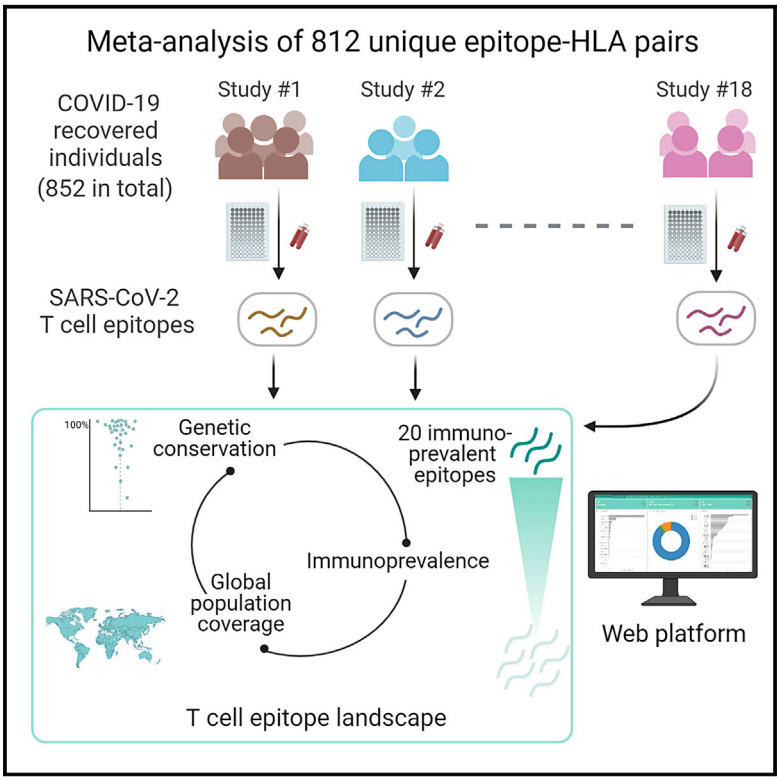
Cell Reports
Medicine

Report

Landscape of epitopes targeted by T cells in 852 individuals recovered from COVID-19: Meta-analysis, immunoprevalence, and web platform

Graphical abstract



Highlights

- Meta-analysis of T cell epitopes from 18 studies of individuals recovered from COVID-19
- 20 immunoprevalent SARS-CoV-2 T cell epitopes identified across multiple cohorts
- Large majority of epitopes appear unaffected by current variants of concern
- Web dashboard for reporting and analyzing SARS-CoV-2 T cell epitope data

Authors

Ahmed Abdul Quadeer,
Syed Faraz Ahmed, Matthew R. McKay

Correspondence

eeaaquadeer@ust.hk (A.A.Q.),
m.mckay@ust.hk (M.R.M.)

In brief

Quadeer et al. provide a meta-analysis of SARS-CoV-2 T cell epitope data from 18 studies involving 852 individuals recovered from COVID-19. Their analysis highlights the characteristics of the reported epitopes and identifies 20 immunoprevalent epitopes targeted in multiple cohorts and in a majority of tested individuals. An associated web-platform is reported.

Quadeer et al., 2021, Cell Reports Medicine 2, 100312
June 15, 2021
https://doi.org/10.1016/j.xcrm.2021.100312

CellPress

Cell Reports Medicine

CellPress
OPEN ACCESS

Report

Landscape of epitopes targeted by T cells in 852 individuals recovered from COVID-19: Meta-analysis, immunoprevalence, and web platform

Ahmed Abdul Quadeer,[1,3,*] Syed Faraz Ahmed,[1] and Matthew R. McKay[1,2,4,*]

[1]Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China

[2]Department of Chemical and Biological Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China

[3]Twitter: @ahmedaquadeer

[4]Lead contact

*Correspondence: eeaaquadeer@ust.hk (A.A.Q.), m.mckay@ust.hk (M.R.M.)

https://doi.org/10.1016/j.xcrm.2021.100312

SUMMARY

Knowledge of the epitopes of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) targeted by T cells in recovered (convalescent) individuals is important for understanding T cell immunity against coronavirus disease 2019 (COVID-19). This information can aid development and assessment of COVID-19 vaccines and inform novel diagnostic technologies. Here, we provide a unified description and meta-analysis of SARS-CoV-2 T cell epitopes compiled from 18 studies of cohorts of individuals recovered from COVID-19 (852 individuals in total). Our analysis demonstrates the broad diversity of T cell epitopes that have been recorded for SARS-CoV-2. A large majority are seemingly unaffected by current variants of concern. We identify a set of 20 immunoprevalent epitopes that induced T cell responses in multiple cohorts and in a large fraction of tested individuals. The landscape of SARS-CoV-2 T cell epitopes we describe can help guide immunological studies, including those related to vaccines and diagnostics. A web-based platform has been developed to help complement these efforts.

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of coronavirus disease 2019 (COVID-19), has led to a global public health crisis. Development of COVID-19 vaccines and diagnostic tests are aided by an understanding of the natural protective immune responses to SARS-CoV-2. This includes humoral and cellular immunity, mediated by antibodies and T cells, respectively. A significant amount of COVID-19 research has focused on understanding antibody responses,[1–3] but studies informing on the role of T cells have also started to emerge.

Initial results suggest a potential key role of T cells in protecting against COVID-19.[4,5] Studies of individuals recovered from COVID-19 have detected SARS-CoV-2-specific T cells 9 months after infection,[6] showing promising signs for the potential of T cells to provide lasting immunity. Longevity may be a concern for antibody responses, which have been reported to decline within few months after infection.[7–9] Observations consistent with this have also been reported for the most closely related human CoV, SARS-CoV, for which T cells have been shown to persist up to 17 years after infection,[10] whereas antibody responses waned after a few years.[11]

Characterizing SARS-CoV-2 T cell epitopes as well as their human leukocyte antigen (HLA) association is important for multiple reasons. It informs on the expected SARS-CoV-2 natural or vaccine-induced T cell responses in a population of specific ethnicity or a specific geographical region, which is tied to the composition of HLA alleles prevalent in that population. It can help with assessing the T cell responses that may be induced by COVID-19 vaccines (which currently mainly focus on the spike [S] protein[12]) and provide possible directions for boosting the T cell response by including specific immunodominant epitopes. It can guide vaccine assessment studies probing whether a vaccine induces T cell responses similar to those commonly generated during natural infection. It can also aid with monitoring potential viral escape from T cell responses via genetic mutations and can facilitate development of T cell-based diagnostics for distinguishing recovered from unexposed individuals. T cell-based diagnostics may have advantages over serological assays, given the uncertainties related to the appearance and persistence of SARS-CoV-2-specific antibody responses in infected individuals.[7–9,13]

Here we present a unified account of the current knowledge (as of April 20, 2021) of SARS-CoV-2 T cell epitopes associated with individuals recovered from COVID-19. We collate and analyze data of T cell epitopes that have been identified experimentally in independent studies of blood samples from different cohorts. These data are compiled from 18 studies (15 published,

3 preprints) of T cell responses in a total of 852 individuals recovered from COVID-19 (Table 1). Our analysis highlights the different characteristics of epitopes reported for SARS-CoV-2 and identifies a specific set of epitopes that appear to induce T cell responses in multiple cohorts and in a large fraction of tested individuals. This information regarding SARS-CoV-2 T cell epitopes can provide directions for future immunological studies. We report a web dashboard we developed to support these ongoing scientific efforts.

RESULTS AND DISCUSSION

In the 18 studies we considered, the set of recovered individuals covers a population across four continents and well-distributed age, gender, disease severity, and blood collection time within and across studies (Table 1). Half of the studies have characterized T cell responses against the whole proteome, whereas the others have focused on responses mounted against subsets of SARS-CoV-2 proteins, typically involving the S and nucleocapsid [N] proteins. In the majority of cases, the T cell response was measured in blood samples of individuals against a set of peptides predicted by bioinformatics tools or earlier bioinformatics-based analyses[16,17,32] (reviewed in Sohail et al.[46]), whereas a few studies employed overlapping peptide pools spanning the SARS-CoV-2 proteome. All 18 studies reported optimal epitopes along with cognate HLA information. Of these, 10 studies (1–10 in Table 1) experimentally determined HLA restrictions of the reported epitopes by using multimer qualitative binding. Of the remaining eight studies, one study (11 in Table 1) employed mono-allelic cell line assays to identify specific HLA-restricted epitopes, whereas others (12–18 in Table 1) inferred HLA restrictions using standard functional assays (such as activation-induced markers [AIMs] or intracellular cytokine staining [ICS]/enzyme-linked immunospot [ELISPOT] interferon γ [IFN-γ] release) and HLA haplotype information for individuals. Although the latter set of studies involved some predictive element in deconvolving the HLA allele responsible for the observed response from the individual's haplotype, the reported HLA associations were supported by HLA binding assays or use of accurate peptide-HLA binding prediction methods.

A total of 711 unique T cell epitopes with information of a cognate HLA allele have been reported in the 18 immunological studies we considered (see STAR Methods for details). Of these, 635 are $CD8^+$ (HLA class I restricted), and 76 are $CD4^+$ (HLA class II restricted) (Figure 1A). The set of epitopes covers each protein from the canonical reading frames of SARS-CoV-2 except open reading frame 7b [ORF7b] and ORF10. The largest numbers of epitopes fall within S and ORF1a—the most exposed protein and the longest SARS-CoV-2 ORF, respectively. These epitopes broadly cover almost the entire S protein, including the receptor-binding domain, whereas the majority of epitopes in ORF1a lie within the non-structural protein [nsp3] (PL2-PROPapain-like proteinase) and nsp4 proteins (File S1). These nsp3 and nsp4 proteins participate in assembly of virally induced cytoplasmic double-membrane vesicles, which are necessary for viral replication.[47]

All identified epitopes have high genetic conservation (>0.9) among the ~860,000 SARS-CoV-2 sequences (as of April 20, 2021), except for 24 epitopes (Figure 1B). Of these, three epitopes ($_{612}$YQDVNCTEV$_{620}$ in S and $_{305}$NVLFSTVFPPTSFGP$_{319}$ and $_{309}$STVFPPTSF$_{317}$, in ORF1b) have very low conservation (~0.02) because they encompass mutations (S:D614G and ORF1b:P314L, underlined) that are now dominant globally, with the former mutation reported to have increased virus infectivity and transmission.[48–50] Two-thirds of the 24 epitopes (16 of 24) with low genetic conservation (<0.9) belong to S. These make up ~8% (16 of 208) of all unique T cell epitopes derived from S—4-fold higher than the fraction of epitopes with low conservation derived from all other proteins (~2%, 8 of 503). This can be of interest in the context of T cell responses against current COVID-19 vaccines that focus largely on the S protein, assuming that the T cell epitopes targeted in response to vaccination are similar to those seen in recovered individuals. Although limited data are currently available about epitope-specific T cell responses following vaccination, this assumption is partially supported by a study of eight S-derived epitopes reported to elicit T cell responses in vaccinated persons,[51] of which we observe seven to be targeted in recovered individuals (File S1). Of the 16 S-derived T cell epitopes with low genetic conservation (<0.9), all but one encompass sites harboring mutations that define the variants of concern (VOCs) (Table S1).[52,53] For example, the epitope $_{495}$YGFQPTNGV$_{503}$ encompasses the N501Y mutation associated with the three VOCs B.1.1.7, B.1.351, and P.1, and the epitopes $_{142}$GVYYHKNNK$_{150}$ and $_{144}$YYHKNNKSW$_{152}$ encompass the deletion at Y144 associated with two variants, B.1.1.7 and B.1.525. The higher genetic variation observed in S as opposed to other proteins is, at least in part, likely to be driven by escape from neutralizing antibodies.[54] Despite the potential of S-derived epitopes to escape T cells elicited to target the non-mutated epitopes (by natural infection or vaccination), responses against the significant majority of S-derived epitopes we studied are not expected to be affected by VOCs (Figure 1B). Moreover, at the level of HLA alleles, for each of the 36 HLA alleles associated with S-derived epitopes, each of them is associated with at least one conserved (>0.9) S-derived epitope (File S1). Collectively, the observed heterogeneity of T cell responses against SARS-CoV-2 provides little evidence to suggest that the observed genetic variation in the S protein may significantly affect T cell immunity, in line with a recent report.[55] Escape from T cell pressure may become an important evolutionary factor when strong and diverse selective pressure is imposed by widespread vaccination, and this requires close monitoring.

The overall set of reported epitopes was associated with 52 HLA class I and II alleles (Figure 1C). Most HLA alleles are associated with multiple epitopes, with each of 21 alleles having an association with 14 epitopes or more. The same epitope may be presented by multiple HLA alleles, as evidenced by studies of the related SARS-CoV[56] as well as other viruses.[57] However, for the majority of SARS-CoV-2 epitopes, only a single associated HLA allele has been reported so far (Figure 1D). This appears in part to be due to the limited number of recovered COVID-19 individuals who have been studied (Table 1). The limited number of associated HLA alleles translates to a median global population coverage estimate per epitope of only 12% (Figure 1E). Thus, investigation of additional HLA alleles associated with the identified SARS-CoV-2 epitopes is required to

Table 1. Summary of immunological studies reporting SARS-CoV-2 T cell epitopes targeted in individuals recovered from COVID-19 (as of April 20, 2021)

No.	Study[a]	Geography (Country)	Total Individuals	Gender (Female/ Male)	Median Age (Range)	Disease Severity[b] (Asymptomatic or Mild)	Disease Severity[b] (Moderate or Severe)	Blood Collection Time (Days)[c]	Initial Peptide Selection Procedure[d]	Proteins	Total Peptides Tested	T cell Assay	Total Distinct Epitopes Identified
1	Saini et al.[14]	Denmark	18	6/12	43.5 (29–82)	7	11	As close as possible to the first positive test	NetMHCpan-4.1	all	2,204	multimer qualitative binding	122
2	Kared et al.[15]	Baltimore/ Washington, USA	30	12/18	42.5 (19–77)	N/A	N/A	27–62 after symptom resolution	Grifoni et al.;[16] Prachar et al.[17]	all	408	multimer qualitative binding	46
3	Schulien et al.[18]	Germany	26	14/12	32.5 (24–56)	26	0	24 (14–70) after symptom onset	ANN-4.0, SARS-CoV epitopes[16]	S, N, M, ORF1ab, ORF3a	66	multimer qualitative binding	37
4	Poran et al.[19]	N/A	3	N/A	N/A	N/A	N/A	N/A	HLAthena	all	23	multimer qualitative binding	11
5	Shomuradova et al.[20]	Moscow, Russia	31	16/15	35 (17–59)	21	10	33 (17–49) after positive test/after disease onset	NetMHCpan-4.0 and identity with SARS-CoV > 60%	S	13	multimer qualitative binding	10
6	Nielsen et al.[21]	Denmark	203	92/111	47 (21–79)	17	186	44 (14–129) after symptom onset	N/A	S, N, M	9	multimer qualitative binding	9
7	Chour et al.[22]	USA	2	0/2	50 (30–70)	0	2	6.5 (2–13) post symptom onset	NetMHC-4.0	S	96	multimer qualitative binding	6
8	Sekine et al.[23]	Sweden	66	14/41	51	40	26	53.5 (IQR: 45.5–61) after symptom onset	NetMHCpan-4.1	all	13	multimer qualitative binding	4
9	Nguyen et al.[24]	Melbourne, Australia	7	1/6	32 (19–74)	6	1	90 (5–145) after disease onset	overlapping peptide pools and 5 N-specific immunogenic peptides[25,18,26,27]	S, N, M	N/A	ICS IFN-γ release; multimer qualitative binding	3
10	Rha et al.[28]	South Korea	37	18/19	46 (21–83)	23	14	46 (19–125) after symptom onset	N/A	S, N, M	8	multimer qualitative binding	2
11	Ferretti et al.[26]	New Jersey/ Louisiana, USA	78	55/23	19.5 (0–80)	55	23	49 (11–111) post diagnosis	identification of 20-mer peptides by TScan screen[29] followed by NetMHC-4.0	all	~240	Single-allele ELISA IFN-γ; multimer qualitative binding[e]	28

(*Continued on next page*)

Table 1. *Continued*

No.	Study[a]	Geography (Country)	Total Individuals	Gender (Female/ Male)	Median Age (Range)	Disease Severity[b] (Asymptomatic or Mild)	Disease Severity[b] (Moderate or Severe)	Blood Collection Time (Days)[c]	Initial Peptide Selection Procedure[d]	Proteins	Total Peptides Tested	T cell Assay	Total Distinct Epitopes Identified
12	Nelde et al.[30]	Germany	116	55/61	44 (18–75)	36	80	37.7 (19–52) after positive test	SYFPEITHI-1.0, NetMHCpan-4.0	all	120	ICS IFN-γ release; ELISPOT IFN-γ release	47
13	Hu et al.[31]	Chongqing, China	37	16/21	47 (20–67)	34	3	N/A	NetMHCpan-4.0, SARS-CoV epitopes[32,16]	S, N	78	ELISPOT IFN-γ release	15
14	Habel et al.[25]	Melbourne, Australia	18	10/8	54 (22–76)	11	7	47 (36–102) after disease onset	NetCTLpan, NetMHCpan	S, N, M, ORF1ab	14	ICS IFN-γ release	14
15	Lineburg et al.[33]	Queensland, Australia	37	23/14	51 (20–75)	7	5	62 (46–124) after positive test	overlapping peptide pools followed by peptide matrix analysis	all	N/A	ICS IFN-γ release	4
16	Lee et al.[34]	New South Wales, Australia	2	N/A	N/A	0	2	30–60 after symptom resolution	Network analysis[35] followed by NetMHCpan-4.1 and NetMHCIIpan-4.0	S, N	2	ICS IFN-γ release	2
17	Tarke et al.[36]	San Diego, USA	99	58/41	41 (19–91)	90	9	67 (3–184) after symptom onset	NetMHCpan-4.0	all	7,525	AIM assay[37]	803
18	Peng et al.[27]	UK	42	16/26	57 (20–95)	28	14	42 (30–62) after symptom onset	15- to 18-mer peptides overlapping by 10 residues, SARS-CoV epitopes[32]	all except ORF1	450	ELISPOT IFN-γ release	46[f]

[a]Studies reporting precise epitopes along with the cognate HLA information.
[b]Definition of disease severity varies among studies.
[c]IQR, interquartile range.
[d]NetCTLpan,[38] NetMHC-4.0,[39] NetMHCpan,[40] NetMHCpan-4.0,[41] NetMHCpan-4.1,[42] SYFPEITHI-1.0,[43] ANN-4.0,[44] and HLAthena.[45]
[e]Six (of 28) epitopes were identified using multimer qualitative binding.
[f]Only two (of 46) epitopes that were reported as precise epitopes with the cognate HLA information were considered.

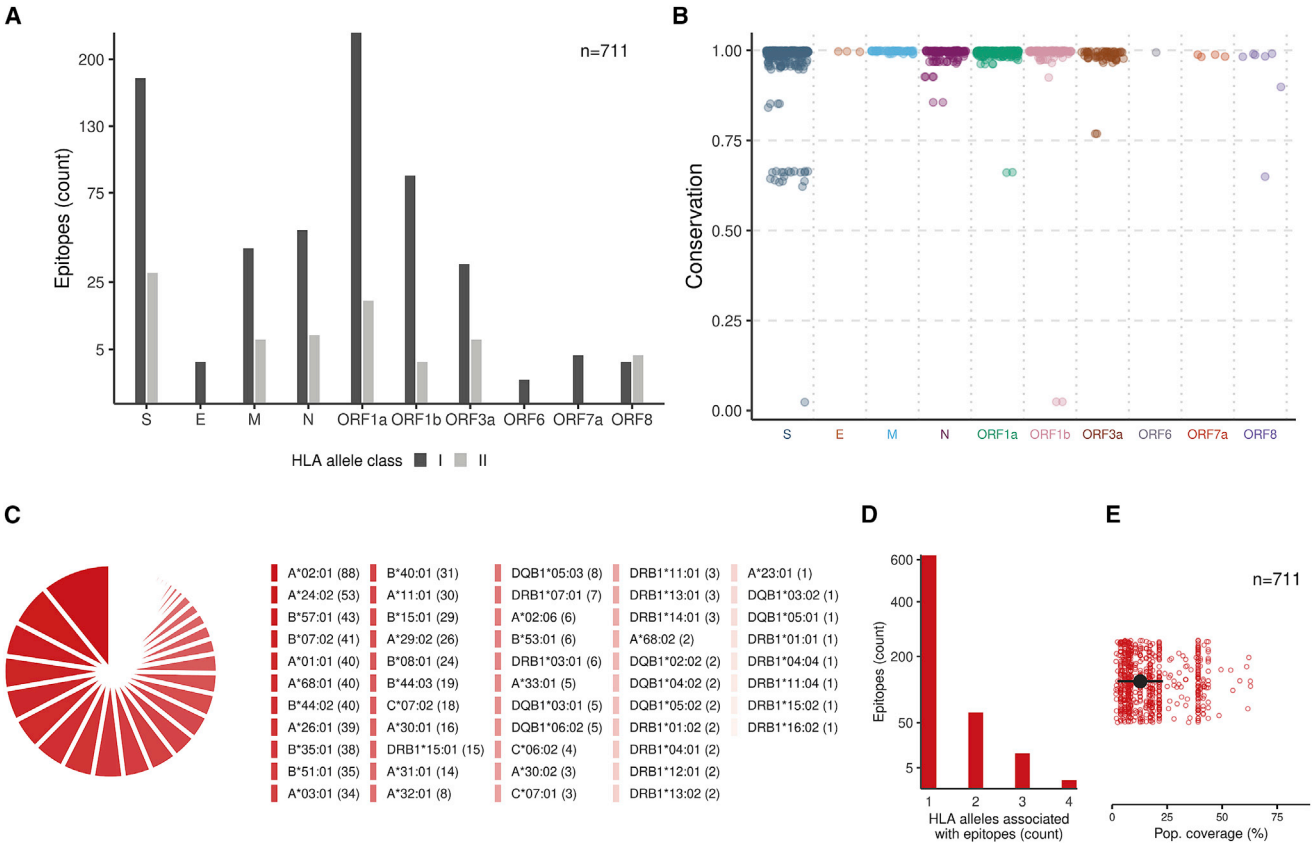


Figure 1. Features of SARS-CoV-2-specific T cell epitopes reported to elicit an immune response in blood samples of individuals recovered from COVID-19

(A) The number of epitopes (n = 711) according to HLA class restriction. S, spike; E, envelope; M, membrane; N, nucleocapsid.

(B) The conservation of each epitope among the global SARS-CoV-2 genetic sequences (as of April 20, 2021). Further details of the S-derived epitopes (n = 16) with low genetic conservation (<0.9) and their association with current VOCs are provided in Table S1.

(C) Diversity of HLA associations reported for SARS-CoV-2 epitopes. The number of epitopes associated with a particular HLA allele is shown in brackets.

(D) The number of HLA alleles associated with each epitope.

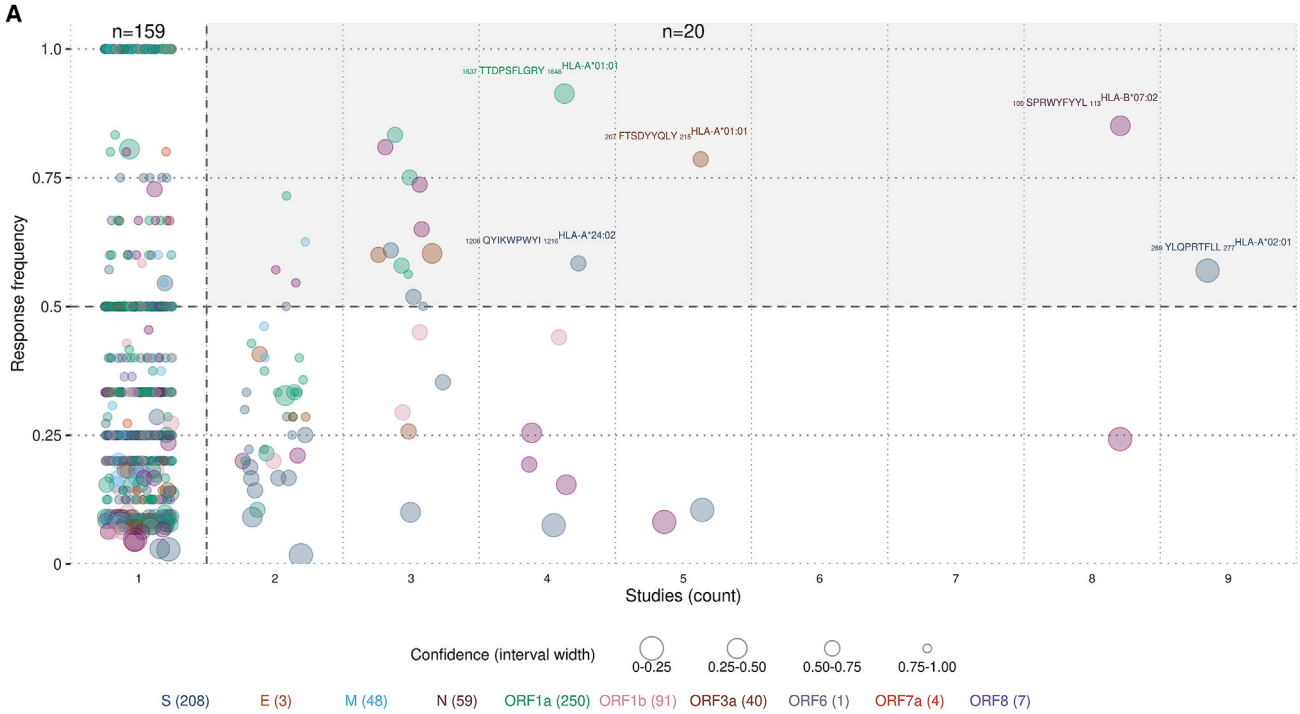
(E) Estimate of the global population coverage of each epitope (STAR Methods). The median is shown as a black circle and the median absolute deviation as an error bar.

providing a more accurate indication of their individual population coverage. An expanded list of likely HLA associations may be predicted for some of the reported epitopes by using prior knowledge of genetically matched experimentally determined T cell epitopes of SARS-CoV and their associated HLA alleles (for details, see File S1 and Figure S1). These predictions, when confirmed, would provide an increase in the median population coverage of the selected SARS-CoV-2 epitopes from 16.8% to 40%, with a few epitopes having around 60% coverage (Figure S1).

Quantifying the immunodominance of each reported epitope-HLA pair using the standard response frequency (RF) metric[58–60] (STAR Methods) revealed that 179 of the reported (812) epitope-HLA pairs had an RF score exceeding 0.5, indicating that they induced T cell responses in over half of the subjects tested across studies (Figure 2A). Confidence in the estimated RF values varies with the number of tested subjects, with higher confidence being attributed to epitope-HLA pairs with larger numbers of tested subjects (STAR Methods). The majority (159 of 179) of epitope-HLA pairs with an RF exceeding 0.5 had been reported in a single study only. Among these, pairs with high confidence appear promising, but responses against them should be investigated in different cohorts to further confirm their immunodominance. Responses against the remaining (20 of 179) epitope-HLA pairs were registered in more than one study, and although per-study variation was observed in the results (Figure S2), these pairs appear to be immunoprevalent.[61] This is because responses against each epitope-HLA pair were recorded for more than half of the tested recovered individuals collectively across multiple studies despite differences in characteristics of the donor cohorts (age, gender, geographical location, and disease severity), blood collection time, and methodology used to determine the epitopes (initial peptide selection procedure and T cell assay) (Table 1).

All of the 20 identified immunoprevalent epitopes have high genetic conservation (>0.9) (Figure 2B), and none of them encompass mutations associated with the current VOCs. Interestingly, 35% of the immunoprevalent epitope-HLA pairs (n = 20) had an

A



B

No.	Epitope	HLA	Protein	Start	End	Conservation	Response Frequency	Studies (References)
1	TTDPSFLGRY	HLA-A*01:01	ORF1a	1637	1646	0.983	0.913	39,47,51,55
2	SPRWYFYYL#	HLA-B*07:02	N	105	113	0.999	0.851	33,34,39,40,41,46,47,55
3	ILFTRFFYV	HLA-A*02:01	ORF1a	2332	2340	0.995	0.833	35,41,55
4	KTFPPTEPK#	HLA-A*11:01	N	361	369	0.969	0.810	40,47,52
5	FTSDYYQLY	HLA-A*01:01	ORF3a	207	215	0.996	0.786	39,40,41,47,55
6	HTTDPSFLGRY	HLA-A*01:01	ORF1a	1636	1646	0.983	0.750	39,40,55
7	ATEGALNTPK#	HLA-A*11:01	N	134	143	0.988	0.737	47,51,55
8	TTDPSFLGRYM	HLA-A*01:01	ORF1a	1637	1647	0.982	0.714	39,55
9	KTFPPTEPK#	HLA-A*03:01	N	361	369	0.969	0.650	39,40,47
10	FLFLTWICL	HLA-A*02:01	M	26	34	0.995	0.625	35,42
11	RLQSLQTYV#	HLA-A*02:01	S	1000	1008	0.999	0.609	42,43,55
12	LLYDANYFL	HLA-A*02:01	ORF3a	139	147	0.995	0.603	40,41,47
13	VYFLQSINF	HLA-A*24:02	ORF3a	112	120	0.994	0.600	40,47,51
14	QYIKWPWYI	HLA-A*24:02	S	1208	1216	0.984	0.583	40,47,51,55
15	PTDNYITTY	HLA-A*01:01	ORF1a	1321	1329	0.996	0.579	39,40,47
16	KTFPPTEPKK#	HLA-A*03:01	N	361	370	0.968	0.571	51,55
17	YLQPRTFLL	HLA-A*02:01	S	269	277	0.955	0.570	33,35,40,43,44,46,47,49,55
18	CTDDNALAYY#	HLA-A*01:01	ORF1a	4163	4172	0.998	0.562	39,41,47
19	KLDDKDPNF	HLA-A*02:01	N	338	346	0.997	0.545	42,52
20	LTDEMIAQY	HLA-A*01:01	S	865	873	0.998	0.519	40,41,51

(legend on next page)

identical match to experimentally determined SARS-CoV epitope-HLA pairs (Figure 2B). This fraction of matched epitope-HLA pairs is over three times higher than that (11%) in the complete set of SARS-CoV-2 epitope-HLA pairs ($p < 10^{-3}$, Fisher's exact test), suggesting that many of the immunoprevalent T cell epitopes of SARS-CoV-2 are also cross-reactive to SARS-CoV. Of the 20 identified immunoprevalent epitope-HLA pairs, six each belonged to N and ORF1a, four to S, three to ORF3a, and one to the membrane [M] protein (Figure 2B). This indicates that ~80% of the immunoprevalent epitopes lie in proteins other than S, suggesting that vaccine candidates targeting these proteins may have benefits in terms of T cell immunity. Of the identified immunoprevalent epitopes, five epitopes (HLA-A*02:01-restricted $_{269}$YLQPRTFLL$_{277}$ in S, HLA-B*07:02-restricted $_{105}$SPRWYFYYL$_{113}$ in N, HLA-A*01:01-restricted $_{207}$FTSDYYQLY$_{215}$ in ORF3a and $_{1637}$TTDPSFLGRY$_{1646}$ in ORF1a, and HLA-A*24:02-restricted $_{1208}$QYIKWPWYI$_{1216}$ in S) appeared to be highly immunoprevalent, eliciting T cell responses in more than ~60% of the tested individuals recovered from COVID-19 in four immunological studies or more. Collectively, around 71% of the global population is estimated to carry the associated HLA alleles and, hence, may generate a T cell response against at least one of these five epitopes. Two of these highly immunoprevalent epitopes ($_{105}$SPRWYFYYL$_{113}$ in N and $_{269}$YLQPRTFLL$_{277}$ in S) are attracting considerable attention, and detailed analyses of T cell responses against them have been reported.[24,25,33]

The SARS-CoV-2 T cell epitope data we compiled and report here were integrated into a web-based dashboard (Figure 3).[62] This dashboard provides exportable data tables listing the SARS-CoV-2 epitopes and graphic displays to summarize different characteristics of the epitopes, including aggregate information as well as specific details of the individual epitopes. We plan to update the dashboard with new experimental information as it becomes available, with the goal of aiding further research to understand T cell responses against SARS-CoV-2 and guide studies related to COVID-19 vaccines and diagnostics. Although we focused the current study on SARS-CoV-2-specific T cell responses recorded in recovered individuals, knowledge of T cell epitopes targeted in animal studies[63,64] is also informative. These may inform potential immune targets in COVID-19-infected individuals and can help guide further immunological experiments that seek to probe T cell responses that arise because of natural infection or those elicited by vaccination. The SARS-CoV-2 epitopes reported to be targeted in animal models were also incorporated into the web dashboard.

Overall, the data we described, based on recent experimental studies, demonstrates an impressive and diverse list of SARS-CoV-2 T cell epitopes targeted by individuals recovered from COVID-19. Subsets of these epitopes exhibit desirable properties, including high genetic conservation and high RF across multiple cohorts, and they appear to have the potential to collectively induce a T cell response in a large fraction of the population. Current knowledge of the landscape of T cell epitopes for SARS-CoV-2 is still evolving, and further studies of different cohorts of recovered individuals, encompassing a broad diversity of HLA profiles, are required to provide a more comprehensive understanding. Moreover, further systematic studies are required to ascertain possible correlates between the responses against T cell epitopes and disease protection. Knowledge of SARS-CoV-2 T cell epitopes could play an important role in contributing to the fight against COVID-19 by guiding diverse applications and novel technologies, including development, assessment, and monitoring of vaccines and development of improved diagnostic assays.

LIMITATIONS OF THE STUDY

There are multiple limitations of our study. Our analysis is unable to make associations between epitope-specific T cell responses and levels of disease severity, nor can it capture differences between T cell responses according to age or gender. The majority of the immunological studies we investigate do not report epitope-specific T cell responses at this level of detail, and, hence, such an analysis could not be performed. In terms of the HLA associations of the reported SARS-CoV-2 epitopes, our study predicted additional associations based on homology with experimental T cell epitopes of SARS-CoV and their associated HLA alleles (File S1). These additional HLA associations, although promising, still need to be confirmed for SARS-CoV-2 by further immunological studies. Moreover, our analysis mainly involved $CD8^+$ T cell epitopes. This was due to the limited number of $CD4^+$ T cell epitopes that have been reported so far with a unique HLA allele association. Further experimental studies are needed to precisely identify $CD4^+$ T cell epitopes with distinct HLA alleles. This would help to determine potential immunoprevalent $CD4^+$ epitopes in the global population, similar to those reported for $CD8^+$ T cells in this work, and would contribute to providing a more comprehensive understanding of the relationship between T cell responses and convalescence for COVID-19.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact

Figure 2. Identification of immunoprevalent SARS-CoV-2 T cell epitopes

(A) Response frequency (RF) of unique epitope-HLA pairs versus the number of immunological studies reporting a T cell response against them. The size of each circle represents the confidence in the respective RF value (STAR Methods). The 20 immunoprevalent epitope-HLA pairs (having an RF exceeding 0.5 and reported in more than one study) are shown with a shaded background, and five highly immunoprevalent epitope-HLA pairs (having an RF exceeding 0.5 and reported in at least four studies) are labeled.

(B) Details of the identified 20 immunoprevalent epitope-HLA pairs (ordered according to decreasing RF). Epitope-HLA pairs matched genetically to those determined experimentally for SARS-CoV are marked (#).

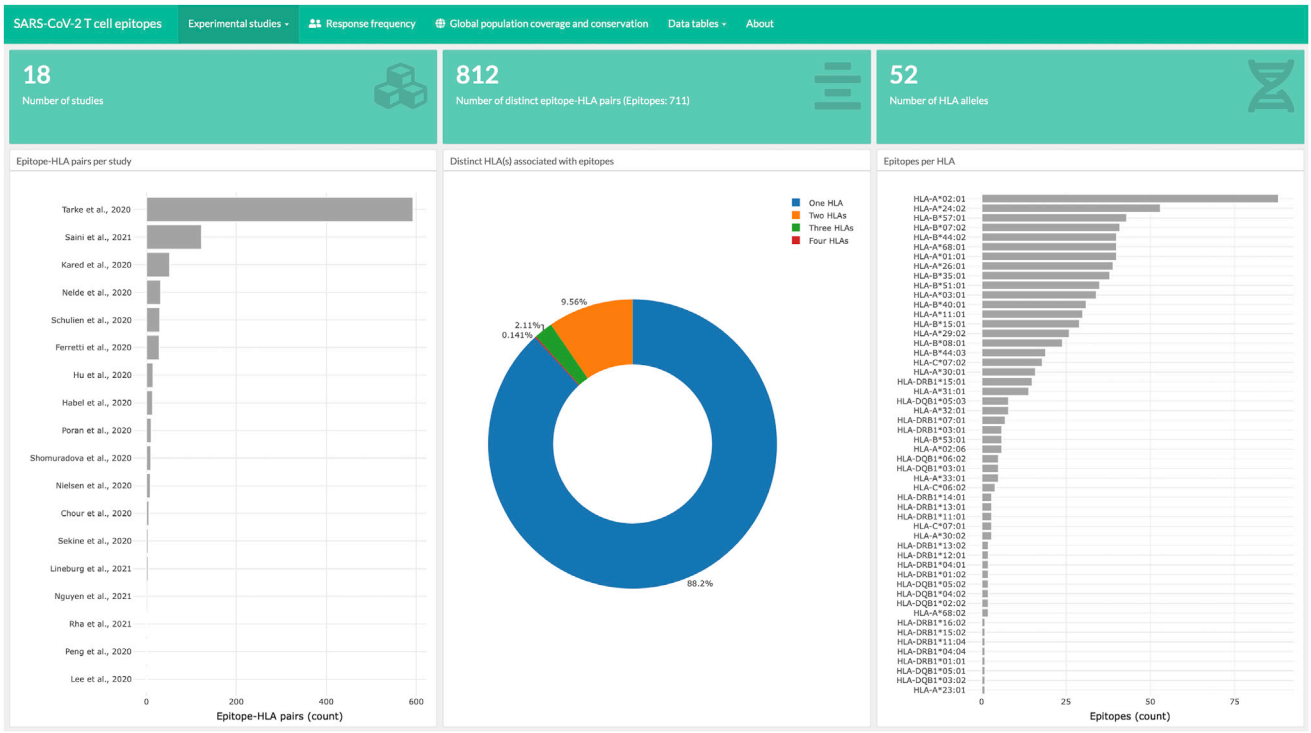


Figure 3. Snapshot of the web dashboard developed for reporting and analyzing SARS-CoV-2 T cell epitope data (as of April 20, 2021)
The web dashboard[62] provides aggregated information regarding the T cell epitopes and their HLA associations. Exportable data tables are provided to aid further research.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xcrm.2021.100312.

ACKNOWLEDGMENTS

The authors were supported by the General Research Fund of the Hong Kong Research Grants Council (RGC) (grants 16204519 and 16201620). S.F.A. was additionally supported by the Hong Kong Ph.D. Fellowship Scheme (HKPFS).

AUTHOR CONTRIBUTIONS

A.A.Q. and M.R.M. conceptualized the study and wrote the manuscript. A.A.Q. compiled the data from the literature. A.A.Q., S.F.A., and M.R.M. analyzed the data. A.A.Q. and S.F.A. performed the computations. S.F.A. generated the figures and developed the web dashboard.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 30, 2020
Revised: January 18, 2021
Accepted: May 14, 2021
Published: May 21, 2021

REFERENCES

1. Yong, S.E.F., Anderson, D.E., Wei, W.E., Pang, J., Chia, W.N., Tan, C.W., Teoh, Y.L., Rajendram, P., Toh, M.P.H.S., Poh, C., et al. (2020). Connecting clusters of COVID-19: an epidemiological and serological investigation. Lancet Infect. Dis. *20*, 809–815.
2. Zohar, T., and Alter, G. (2020). Dissecting antibody-mediated protection against SARS-CoV-2. Nat. Rev. Immunol. *20*, 392–394.
3. Wu, F., Wang, A., Liu, M., Wang, Q., Chen, J., Xia, S., Ling, Y., Zhang, Y., Xun, J., Lu, L., et al. (2020). Neutralizing antibody responses to SARS-CoV-2 in a COVID-19 recovered patient cohort and their implications. medRxiv. https://doi.org/10.1101/2020.03.30.20047365.
4. Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jadi, R.S., et al. (2020). Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. Cell *181*, 1489–1501.e15.
5. Weiskopf, D., Schmitz, K.S., Raadsen, M.P., Grifoni, A., Okba, N.M.A., Endeman, H., van den Akker, J.P.C., Molenkamp, R., Koopmans, M.P.G., van Gorp, E.C.M., et al. (2020). Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. Sci. Immunol. *5*, eabd2071.
6. Li, Z., Liu, J., Deng, H., Yang, X., Wang, H., Feng, X., Zelinskyy, G., Trilling, M., Sutter, K., Lu, M., et al. (2020). SARS-CoV-2-specific T cell memory is

long-lasting in the majority of convalescent COVID-19 individuals. bioRxiv. https://doi.org/10.1101/2020.11.15.383463.

7. Snyder, T.M., Gittelman, R.M., Klinger, M., May, D.H., Osborne, E.J., Taniguchi, R., Zahid, H.J., Kaplan, I.M., Dines, J.N., Noakes, M.N., et al. (2020). Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. medRxiv. https://doi.org/10.1101/2020.07.31.20165647.
8. Long, Q.-X., Tang, X.-J., Shi, Q.-L., Li, Q., Deng, H.-J., Yuan, J., Hu, J.-L., Xu, W., Zhang, Y., Lv, F.-J., et al. (2020). Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. Nat. Med. *26*, 1200–1204.
9. Seow, J., Graham, C., Merrick, B., Acors, S., Steel, K.J.A., Hemmings, O., O'Bryne, A., Kouphou, N., Pickering, S., Galao, R., et al. (2020). Longitudinal evaluation and decline of antibody responses in SARS-CoV-2 infection. medRxiv. https://doi.org/10.1101/2020.07.09.20148429.
10. Le Bert, N., Tan, A.T., Kunasegaran, K., Tham, C.Y.L., Hafezi, M., Chia, A., Chng, M.H.Y., Lin, M., Tan, N., Linster, M., et al. (2020). SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. Nature *584*, 457–462.
11. Tang, F., Quan, Y., Xin, Z.-T., Wrammert, J., Ma, M.-J., Lv, H., Wang, T.-B., Yang, H., Richardus, J.H., Liu, W., and Cao, W.C. (2011). Lack of peripheral memory B cell responses in recovered patients with severe acute respiratory syndrome: a six-year follow-up study. J. Immunol. *186*, 7264–7268.
12. Le, T.T., Cramer, J.P., Chen, R., and Mayhew, S. (2020). Evolution of the COVID-19 vaccine development landscape. Nat. Rev. Drug Discov. *19*, 667–668.
13. Liu, Z.-L., Liu, Y., Wan, L.-G., Xiang, T.-X., Le, A.-P., Liu, P., Peiris, M., Poon, L.L.M., and Zhang, W. (2020). Antibody profiles in mild and severe cases of COVID-19. Clin. Chem. *66*, 1102–1104.
14. Saini, S.K., Hersby, D.S., Tamhane, T., Povlsen, H.R., Amaya Hernandez, S.P., Nielsen, M., Gang, A.O., and Hadrup, S.R. (2021). SARS-CoV-2 genome-wide T cell epitope mapping reveals immunodominance and substantial $CD8^+$ T cell activation in COVID-19 patients. Sci. Immunol. *6*, eabf7550.
15. Kared, H., Redd, A.D., Bloch, E.M., Bonny, T.S., Sumatoh, H., Kairi, F., Carbajo, D., Abel, B., Newell, E.W., Bettinotti, M.P., et al. (2021). SARS-CoV-2-specific CD8+ T cell responses in convalescent COVID-19 individuals. J. Clin. Invest. *131*, e145476.
16. Grifoni, A., Sidney, J., Zhang, Y., Scheuermann, R.H., Peters, B., and Sette, A. (2020). A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2. Cell Host Microbe *27*, 671–680.e2.
17. Prachar, M., Justesen, S., Steen-Jensen, D.B., Thorgrimsen, S., Jurgons, E., Winther, O., and Bagger, F.O. (2020). Identification and validation of 174 COVID-19 vaccine candidate epitopes reveals low performance of common epitope prediction tools. Sci. Rep. *10*, 20465.
18. Schulien, I., Kemming, J., Oberhardt, V., Wild, K., Seidel, L.M., Killmer, S., Sagar, Daul, F., Salvat Lago, M., Decker, A., et al. (2021). Characterization of pre-existing and induced SARS-CoV-2-specific $CD8^+$ T cells. Nat. Med. *27*, 78–85.
19. Poran, A., Harjanto, D., Malloy, M., Arieta, C.M., Rothenberg, D.A., Lenkala, D., van Buuren, M.M., Addona, T.A., Rooney, M.S., Srinivasan, L., and Gaynor, R.B. (2020). Sequence-based prediction of SARS-CoV-2 vaccine targets using a mass spectrometry-based bioinformatics predictor identifies immunogenic T cell epitopes. Genome Med. *12*, 70.
20. Shomuradova, A.S., Vagida, M.S., Sheetikov, S.A., Zornikova, K.V., Kiryukhin, D., Titov, A., Peshkova, I.O., Khmelevskaya, A., Dianov, D.V., Malasheva, M., et al. (2020). SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T cell receptors. Immunity *53*, 1245–1257.e5.
21. Nielsen, S.S.F., Vibholm, L.K., Monrad, I., Olesen, R., Frattari, G.S., Pahus, M.H., Højen, J.F., Gunst, J.D., Erikstrup, C., Holleufer, A., et al. (2020). SARS-CoV-2 elicits robust adaptive immune responses regardless of disease severity. bioRxiv. https://doi.org/10.1101/2020.10.08.331645.
22. Chour, W., Xu, A.M., Ng, A.H.C., Choi, J., Xie, J., Yuan, D., Lee, J.K., Delucia, D.C., Edmark, R., Jones, L., et al. (2020). Shared antigen-specific CD8+ T cell responses against the SARS-COV-2 spike protein in HLA A*02:01 COVID-19 participants. medRxiv. https://doi.org/10.1101/2020.05.04.20085779.
23. Sekine, T., Perez-Potti, A., Rivera-Ballesteros, O., Strålin, K., Gorin, J.-B., Olsson, A., Llewellyn-Lacey, S., Kamal, H., Bogdanovic, G., Muschiol, S., et al.; Karolinska COVID-19 Study Group (2020). Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. Cell *183*, 158–168.e14.
24. Nguyen, T.H.O., Rowntree, L.C., Petersen, J., Chua, B.Y., Hensen, L., Kedzierski, L., van de Sandt, C.E., Chaurasia, P., Tan, H.-X., Habel, J.R., et al. (2021). $CD8^+$ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope display high naive precursor frequency and TCR promiscuity. Immunity *54*, 1066–1082.e5.
25. Habel, J.R., Nguyen, T.H.O., van de Sandt, C.E., Juno, J.A., Chaurasia, P., Wragg, K., Koutsakos, M., Hensen, L., Jia, X., Chua, B., et al. (2020). Suboptimal SARS-CoV-2-specific $CD8^+$ T cell response associated with the prominent HLA-A*02:01 phenotype. Proc. Natl. Acad. Sci. USA *117*, 24384–24391.
26. Ferretti, A.P., Kula, T., Wang, Y., Nguyen, D.M.V., Weinheimer, A., Dunlap, G.S., Xu, Q., Nabilsi, N., Perullo, C.R., Cristofaro, A.W., et al. (2020). Unbiased screens show CD8+ T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that largely reside outside the spike Protein. Immunity *53*, 1095–1107.e3.
27. Peng, Y., Mentzer, A.J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., et al.; Oxford Immunology Network Covid-19 Response T cell Consortium; ISARIC4C Investigators (2020). Broad and strong memory $CD4^+$ and $CD8^+$ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. Nat. Immunol. *21*, 1336–1345.
28. Rha, M.-S., Jeong, H.W., Ko, J.-H., Choi, S.J., Seo, I.-H., Lee, J.S., Sa, M., Kim, A.R., Joo, E.-J., Ahn, J.Y., et al. (2021). PD-1-expressing SARS-CoV-2-specific CD8+ T cells are not exhausted, but functional in patients with COVID-19. Immunity *54*, 44–52.e3.
29. Kula, T., Dezfulian, M.H., Wang, C.I., Abdelfattah, N.S., Hartman, Z.C., Wucherpfennig, K.W., Lyerly, H.K., and Elledge, S.J. (2019). T-Scan: A genome-wide method for the systematic discovery of T cell epitopes. Cell *178*, 1016–1028.e13.
30. Nelde, A., Bilich, T., Heitmann, J.S., Maringer, Y., Salih, H.R., Roerden, M., Lübke, M., Bauer, J., Rieth, J., Wacker, M., et al. (2021). SARS-CoV-2-derived peptides define heterologous and COVID-19-induced T cell recognition. Nat. Immunol. *22*, 74–85.
31. Hu, C., Shen, M., Han, X., Chen, Q., Li, L., Chen, S., Zhang, J., Gao, F., Wang, W., Wang, Y., et al. (2020). Identification of Cross-Reactive CD8+ T Cell Receptors with High Functional Avidity to a SARS-CoV-2 Immunodominant Epitope and Its Natural Mutant Variants. bioRxiv. https://doi.org/10.1101/2020.11.02.364729.
32. Ahmed, S.F., Quadeer, A.A., and McKay, M.R. (2020). Preliminary identification of potential vaccine targets for the COVID-19 coronavirus (SARS-CoV-2) based on SARS-CoV immunological studies. Viruses *12*, 254.
33. Lineburg, K.E., Grant, E.J., Swaminathan, S., Chatzileontiadou, D.S.M., Szeto, C., Sloane, H., Panikkar, A., Raju, J., Crooks, P., Rehan, S., et al. (2021). $CD8^+$ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses. Immunity *54*, 1055–1065.e5.
34. Lee, E., Sandgren, K., Duette, G., Stylianou, V.V., Khanna, R., Eden, J.-S., Blyth, E., Gottlieb, D., Cunningham, A.L., and Palmer, S. (2020). Identification of SARS-CoV-2 nucleocapsid and spike T-cell epitopes for assessing T-cell immunity. J. Virol. *95*, e02002-20.
35. Gaiha, G.D., Rossin, E.J., Urbach, J., Landeros, C., Collins, D.R., Nwonu, C., Muzhingi, I., Anahtar, M.N., Waring, O.M., Piechocka-Trocha, A., et al.

(2019). Structural topology defines protective CD8+ T cell epitopes in the HIV proteome. Science *364*, 480–484.

36. Tarke, A., Sidney, J., Kidd, C.K., Dan, J.M., Ramirez, S.I., Yu, E.D., Mateus, J., da Silva Antunes, R., Moore, E., Rubiro, P., et al. (2021). Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. Cell Rep Med *2*, 100204.
37. Reiss, S., Baxter, A.E., Cirelli, K.M., Dan, J.M., Morou, A., Daigneault, A., Brassard, N., Silvestri, G., Routy, J.-P., Havenar-Daughton, C., et al. (2017). Comparative analysis of activation induced marker (AIM) assays for sensitive identification of antigen-specific CD4 T cells. PLoS ONE *12*, e0186998.
38. Stranzl, T., Larsen, M.V., Lundegaard, C., and Nielsen, M. (2010). NetCTLpan: pan-specific MHC class I pathway epitope predictions. Immunogenetics *62*, 357–368.
39. Andreatta, M., and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics *32*, 511–517.
40. Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., and Buus, S. (2007). NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. PLoS ONE *2*, e796.
41. Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: Improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. J. Immunol. *199*, 3360–3368.
42. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. Nucleic Acids Res. *48* (*W1*), W449–W454.
43. Rammensee, H., Bachmann, J., Emmerich, N.P.N., Bachor, O.A., and Stevanović, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. Immunogenetics *50*, 213–219.
44. Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S.L., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. Protein Sci. *12*, 1007–1017.
45. Abelin, J.G., Keskin, D.B., Sarkizova, S., Hartigan, C.R., Zhang, W., Sidney, J., Stevens, J., Lane, W., Zhang, G.L., Eisenhaure, T.M., et al. (2017). Mass spectrometry profiling of HLA-associated peptidomes in mono-allelic cells enables more accurate epitope prediction. Immunity *46*, 315–326.
46. Sohail, M.S., Ahmed, S.F., Quadeer, A.A., and McKay, M.R. (2021). In silico T cell epitope identification for SARS-CoV-2: Progress and perspectives. Adv. Drug Deliv. Rev. *171*, 29–47.
47. UniProt (2020). SARS-CoV-2 replicase polyprotein 1ab. https://covid-19.uniprot.org/uniprotkb/P0DTD1.
48. Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X., Muruato, A.E., Zou, J., Fontes-Garfias, C.R., et al. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. Nature *592*, 116–121.
49. Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., et al.; Sheffield COVID-19 Genomics Group (2020). Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell *182*, 812–827.e19.
50. Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K.H., 3rd, Leist, S.R., Schäfer, A., Nakajima, N., Takahashi, K., et al. (2020). SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science *370*, 1464–1468.
51. Sahin, U. (2020). BNT162b2 induces SARS-CoV-2-neutralising antibodies and T cells in humans. medRxiv, 2020.12.09.20245175.
52. Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., et al. (2021). Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature *593*, 130–135.
53. Wang, P., Casner, R.G., Nair, M.S., Wang, M., Yu, J., Cerutti, G., Liu, L., Kwong, P.D., Huang, Y., Shapiro, L., and Ho, D.D. (2021). Increased resistance of SARS-CoV-2 variant P.1 to antibody neutralization. Cell Host Microbe *29*, 747–751.e4.
54. Greaney, A.J., Starr, T.N., Gilchuk, P., Zost, S.J., Binshtein, E., Loes, A.N., Hilton, S.K., Huddleston, J., Eguia, R., Crawford, K.H.D., et al. (2021). Complete mapping of mutations to the SARS-CoV-2 spike receptor-binding domain that escape antibody recognition. Cell Host Microbe *29*, 44–57.e9.
55. Tarke, A., Sidney, J., Methot, N., Zhang, Y., Dan, J.M., Goodwin, B., Rubiro, P., Sutherland, A., da Silva Antunes, R., Frazier, A., et al. (2021). Negligible impact of SARS-CoV-2 variants on CD4+ and CD8+ T cell reactivity in COVID-19 exposed donors and vaccinees. bioRxiv. https://doi.org/10.1101/2021.02.27.433180.
56. Ahmed, S.F., Quadeer, A.A., and McKay, M.R. (2020). COVIDep: a web-based platform for real-time reporting of vaccine target recommendations for SARS-CoV-2. Nat. Protoc. *15*, 2141–2142.
57. Vita, R., Mahajan, S., Overton, J.A., Dhanda, S.K., Martini, S., Cantrell, J.R., Wheeler, D.K., Sette, A., and Peters, B. (2019). The Immune Epitope Database (IEDB): 2018 update. Nucleic Acids Res. *47* (*D1*), D339–D343.
58. Dhanda, S.K., Vita, R., Ha, B., Grifoni, A., Peters, B., and Sette, A. (2018). ImmunomeBrowser: a tool to aggregate and visualize complex and heterogeneous epitopes in reference proteins. Bioinformatics *34*, 3931–3933.
59. Tian, Y., Grifoni, A., Sette, A., and Weiskopf, D. (2019). Human T cell response to dengue virus infection. Front. Immunol. *10*, 2125.
60. Weiskopf, D., Angelo, M.A., de Azeredo, E.L., Sidney, J., Greenbaum, J.A., Fernando, A.N., Broadwater, A., Kolla, R.V., De Silva, A.D., de Silva, A.M., et al. (2013). Comprehensive analysis of dengue virus-specific responses supports an HLA-linked protective role for CD8+ T cells. Proc. Natl. Acad. Sci. USA *110*, E2046–E2053.
61. Sidney, J., Peters, B., and Sette, A. (2020). Epitope prediction and identification- adaptive T cell responses in humans. Semin. Immunol. *50*, 101418.
62. Ahmed, S.F., Quadeer, A.A., and McKay, M.R. (2020). A web-based dashboard for reporting epitopes targeted by T cells in convalescent COVID-19 patients. https://www.mckayspcb.com/SARS2TcellEpitopes/.
63. Corbett, K.S., Edwards, D.K., Leist, S.R., Abiona, O.M., Boyoglu-Barnum, S., Gillespie, R.A., Himansu, S., Schäfer, A., Ziwawo, C.T., DiPiazza, A.T., et al. (2020). SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. Nature *586*, 567–571.
64. Smith, T.R.F., Patel, A., Ramos, S., Elwood, D., Zhu, X., Yan, J., Gary, E.N., Walker, S.N., Schultheis, K., Purwar, M., et al. (2020). Immunogenicity of a DNA vaccine candidate for COVID-19. Nat. Commun. *11*, 2601.
65. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
T cell epitopes from convalescent COVID-19 patients	Chour et al.[22]	Figure 3
T cell epitopes from convalescent COVID-19 patients	Shomuradova et al.[20]	Table 1
T cell epitopes from convalescent COVID-19 patients	Nelde et al.[30]	Extended Data Tables 2 and 3
T cell epitopes from convalescent COVID-19 patients	Poran et al.[19]	Figure 3
T cell epitopes from convalescent COVID-19 patients	Ferretti et al.[26]	Table 1
T cell epitopes from convalescent COVID-19 patients	Peng et al.[27]	Tables 1 and 2
T cell epitopes from convalescent COVID-19 patients	Kared et al.[15]	Figure S2
T cell epitopes from convalescent COVID-19 patients	Schulien et al.[18]	Table S1
T cell epitopes from convalescent COVID-19 patients	Habel et al.[25]	Figure 2B
T cell epitopes from convalescent COVID-19 patients	Hu et al.[31]	Table 1
T cell epitopes from convalescent COVID-19 patients	Nielsen et al.[21]	Figure 5B
T cell epitopes from convalescent COVID-19 patients	Tarke et al.[36]	Tables S3 and S5
T cell epitopes from convalescent COVID-19 patients	Sekine et al.[23]	Table S2
T cell epitopes from convalescent COVID-19 patients	Saini et al.[14]	Table S5
T cell epitopes from convalescent COVID-19 patients	Lee et al.[34]	Figure 13
T cell epitopes from convalescent COVID-19 patients	Rha et al.[28]	Figure 1C
T cell epitopes from convalescent COVID-19 patients	Lineburg et al.[33]	Table S3
T cell epitopes from convalescent COVID-19 patients	Nguyen et al.[24]	Figure 3C
Response frequencies of SARS-CoV-2 T cell epitopes	This paper	File S1; Mendeley data: https://dx.doi.org/10.17632/fwn3kbbh6y.1
Genetic conservation of SARS-CoV-2 T cell epitopes among SARS-CoV-2 sequences (as of 20 April 2021)	This paper	File S1; Mendeley data: https://dx.doi.org/10.17632/fwn3kbbh6y.1
Total subjects tested for each SARS-CoV-2 T cell epitope across studies	This paper	File S1; Mendeley data: https://dx.doi.org/10.17632/fwn3kbbh6y.1
Total subjects responded to each SARS-CoV-2 T cell epitope across studies	This paper	File S1; Mendeley data: https://dx.doi.org/10.17632/fwn3kbbh6y.1
Additional HLA alleles predicted for SARS-CoV-2 T cell epitopes	This paper	File S1; Mendeley data: https://dx.doi.org/10.17632/fwn3kbbh6y.1

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SARS-CoV-2 genome sequence data for conservation analysis	GISAID (https://www.gisaid.org)	All full genome and high coverage sequences available as of 20 April 2021
SARS-CoV-2 reference genome for aligning the sequences	https://www.ncbi.nlm.nih.gov/nuccore/NC_045512.2	GenBank: NC_045512.2
SARS-CoV T cell epitope data	IEDB; https://www.iedb.org/	IEDB was queried for epitopes with positive MHC binding and positive T cell assays using "Severe acute respiratory syndrome-related coronavirus" as "Organism" on 21 February 2020.
Software and algorithms		
SARS-CoV-2 T cell epitope web-dashboard	This paper	https://www.mckayspcb.com/SARS2TcellEpitopes/
MAFFT	Katoh and Standley[65]	https://mafft.cbrc.jp/alignment/software/
Estimated population coverage of SARS-CoV-2 T cell epitopes based on associated HLA alleles	IEDB Analysis Resource - Population coverage tool	http://tools.iedb.org/population/download/

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact Matthew R. McKay (m.mckay@ust.hk).

Materials availability

This study did not generate new materials.

Data and code availability

Compiled data of the SARS-CoV-2 T cell epitopes is available to download from the web dashboard, https://www.mckayspcb.com/SARS2TcellEpitopes/. File S1 has been deposited to Mendeley Data: http://dx.doi.org/10.17632/fwn3kbbh6y.1.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

We compiled data from 20 immunological studies that reported SARS-CoV-2 epitopes targeted by T cells in individuals recovered from COVID-19. Two studies[7,10] reported responses against synthetic peptide pool libraries using functional or molecular assays, and identified long immunogenic peptides. As complete information of epitopes was not available in these studies, we focused on the remaining 18 studies that provided precise epitopes along with their HLA restriction (Table 1). All statistics related to the patients that participated in each of these 18 studies (age, gender, geographical location, disease severity, blood collection time) are summarized in Table 1. Across the considered 18 studies, the total number of recovered COVID-19 individuals was 852. A total of 1,209 epitopes were obtained from these immunological studies (Table 1). Removing the epitopes with no HLA allele information and those for which the number of tested and responded patients was not reported at a distinct epitope-HLA level resulted in a total of 711 unique epitopes (812 unique epitope-HLA pairs; listed in File S1).

METHOD DETAILS

Sequence data, epitope conservation and coverage

SARS-CoV-2 genomic sequences were obtained from the GISAID database (https://www.gisaid.org/) on 20 April 2021. We downloaded only the complete (full-genome) sequences derived from human hosts with high coverage using the options provided on the GISAID database. All of the 859,233 downloaded sequences were aligned to the SARS-CoV-2 reference genome (GenBank: NC_045512.2) using MAFFT.[65] The genomic MSA was translated using an in-house code to obtain the protein MSAs. The positions of the open reading frames provided with the reference sequence were used to identify the respective protein regions of the full genome.

The conservation of each SARS-CoV-2 T cell epitope was calculated as the fraction of SARS-CoV-2 sequences that encompassed the precise epitope sequence. The coverage of SARS-CoV-2 T cell epitopes at any position of a protein was calculated by counting the number of epitopes that included that position.

Response frequency (RF)

RF score[58] was used to quantify the immunodominance of the SARS-CoV-2 epitopes reported to be recognized by T cells in recovered COVID-19 individuals. The RF score of an epitope is defined as follows:

$$RF = \frac{\sum_{i=1}^{S} r_i}{\sum_{i=1}^{S} t_i}$$

where r_i is the number of subjects responding to the epitope in study i, t_i is the number of subjects tested for a response against the epitope in study i, and S is the total number of studies. An RF score calculated using a large number of tested subjects would be more reliable than one calculated using relatively few subjects. To account for this, we computed the 95% confidence interval for the RF score of each epitope using the binomial cumulative distribution function[58]. In Figure 2, we defined the confidence in RF value of an epitope as the inverse of the length of the corresponding 95% confidence interval. That is, values of RF with a smaller 95% confidence interval have higher confidence, and vice versa.

Estimating global population coverage of epitopes

The global population coverage of an epitope refers to the percentage of individuals in the world population that is expected to mount a T cell response against that epitope. The population coverage of a T cell epitope was calculated based on the HLA alleles associated with it using the tool downloaded from the IEDB Analysis Resource (http://tools.iedb.org/population/download/). This tool employs global HLA allele frequency data obtained from the Allele Frequency Net Database (http://www.allelefrequencies.net/) to estimate the population coverage.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analyses were performed using MATLAB (R2019b) and the R language (version 3.6) using the RStudio server (version 1.3). The web-based platform was developed using the open source R Shiny (version 1.5) development framework. Fisher's exact test was used to compute the statistical significance associated with enrichment of SARS-CoV epitopes among immunoprevalent SARS-CoV-2 epitopes. The 95% confidence interval for the RF score of each epitope in Figure 2 was determined using the binomial cumulative distribution function.[58]