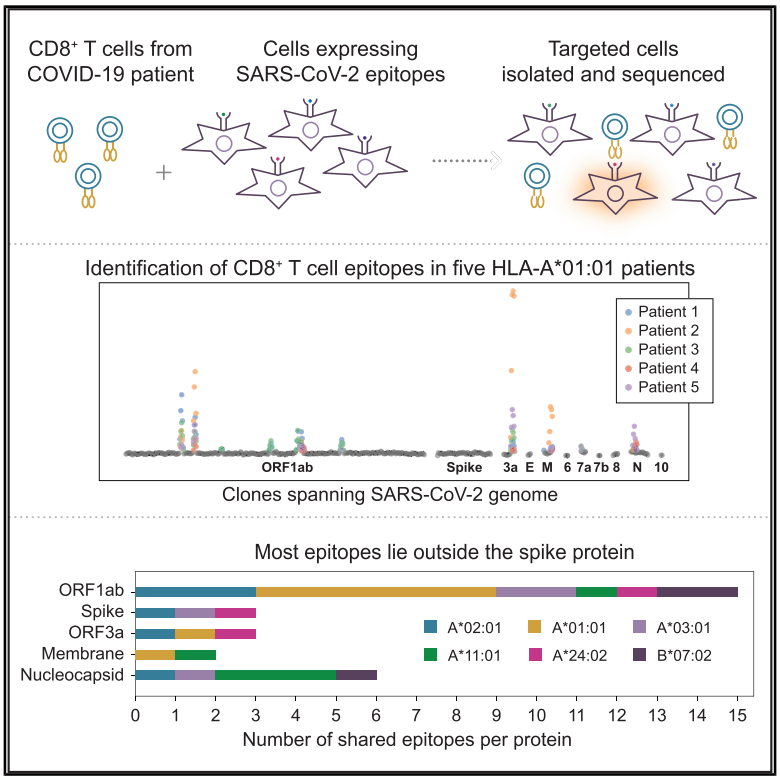
Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Article

Immunity

Unbiased Screens Show CD8+ T Cells of COVID-19 Patients Recognize Shared Epitopes in SARS-CoV-2 that Largely Reside outside the Spike Protein

Graphical Abstract



Highlights

- Unbiased screens identified SARS-CoV-2 targets of CD8+ T cells in COVID-19 patients
- CD8+ T cells predominantly recognize 3–8 shared epitopes for each HLA type studied
- ~90% of shared epitopes are not located in the spike protein
- CD8+ T cells show almost no cross-reactivity with epitopes in seasonal coronaviruses

Authors

Andrew P. Ferretti, Tomasz Kula, Yifan Wang, ..., Sarah A. Bertino, Shrikanta Chattopadhyay, Gavin MacBeath

Correspondence

gmacbeath@tscan.com

In Brief

Ferretti et al. reveal specific SARS-CoV-2 epitopes that are broadly shared by CD8+ T cells of COVID-19 patients but exhibit limited cross-reactivity with seasonal coronaviruses. Most epitopes are located outside of the spike protein, suggesting that next-generation vaccines incorporating these epitopes might be needed to generate more robust and durable immunity.

Ferretti et al., 2020, Immunity *53*, 1095–1107
November 17, 2020
https://doi.org/10.1016/j.immuni.2020.10.006

CellPress

Immunity

CellPress

Article

Unbiased Screens Show CD8+ T Cells of COVID-19 Patients Recognize Shared Epitopes in SARS-CoV-2 that Largely Reside outside the Spike Protein

Andrew P. Ferretti,[1,5] Tomasz Kula,[1,4,5] Yifan Wang,[1] Dalena M.V. Nguyen,[1] Adam Weinheimer,[1] Garrett S. Dunlap,[1] Qikai Xu,[1] Nancy Nabilsi,[1] Candace R. Perullo,[1] Alexander W. Cristofaro,[1] Holly J. Whitton,[1] Amy Virbasius,[1] Kenneth J. Olivier, Jr.,[1] Lyndsey R. Buckner,[2] Angela T. Alistar,[3] Eric D. Whitman,[3] Sarah A. Bertino,[1] Shrikanta Chattopadhyay,[1] and Gavin MacBeath[1,6,*]

[1]TScan Therapeutics, Waltham, MA 02451, USA
[2]Ochsner Medical Center, New Orleans, LA 70121, USA
[3]Atlantic Health System, Morristown, NJ 07960, USA
[4]Present address: Society of Fellows, Harvard University, Cambridge, MA 02138, USA
[5]These authors contributed equally
[6]Lead Contact
*Correspondence: gmacbeath@tscan.com
https://doi.org/10.1016/j.immuni.2020.10.006

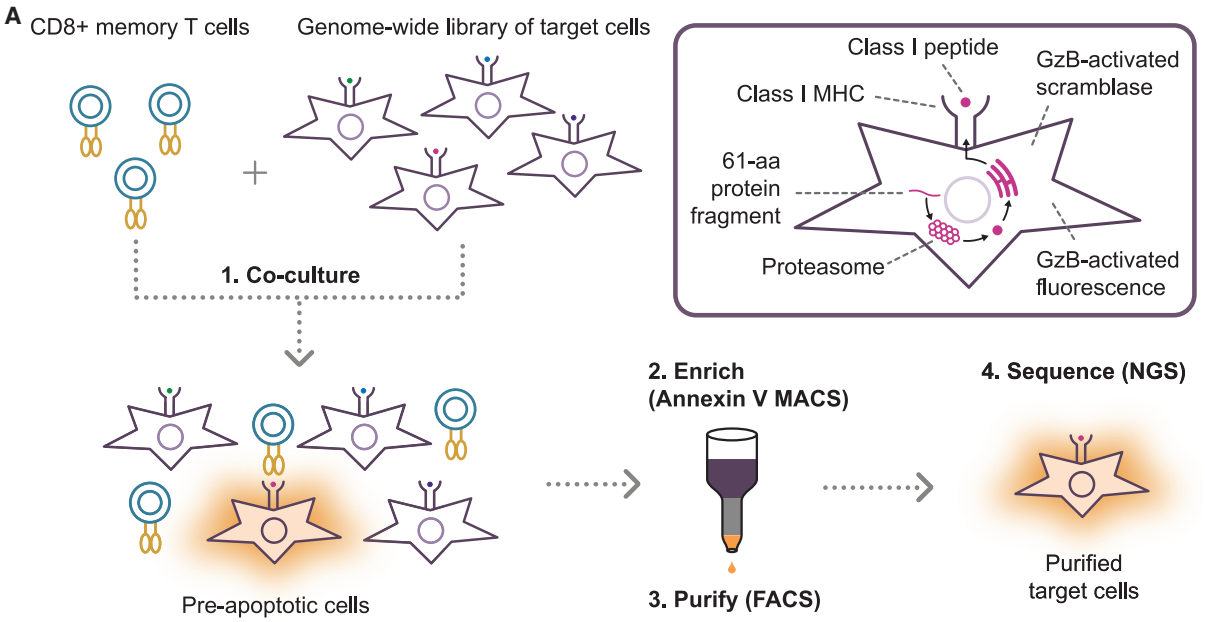
SUMMARY

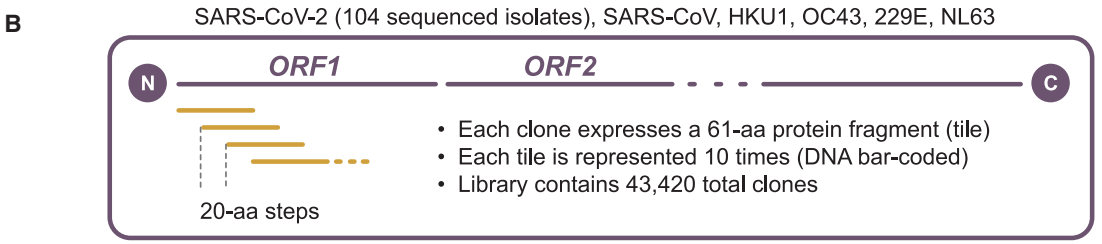
Developing effective strategies to prevent or treat coronavirus disease 2019 (COVID-19) requires understanding the natural immune response to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). We used an unbiased, genome-wide screening technology to determine the precise peptide sequences in SARS-CoV-2 that are recognized by the memory CD8+ T cells of COVID-19 patients. In total, we identified 3–8 epitopes for each of the 6 most prevalent human leukocyte antigen (HLA) types. These epitopes were broadly shared across patients and located in regions of the virus that are not subject to mutational variation. Notably, only 3 of the 29 shared epitopes were located in the spike protein, whereas most epitopes were located in ORF1ab or the nucleocapsid protein. We also found that CD8+ T cells generally do not cross-react with epitopes in the four seasonal coronaviruses that cause the common cold. Overall, these findings can inform development of next-generation vaccines that better recapitulate natural CD8+ T cell immunity to SARS-CoV-2.

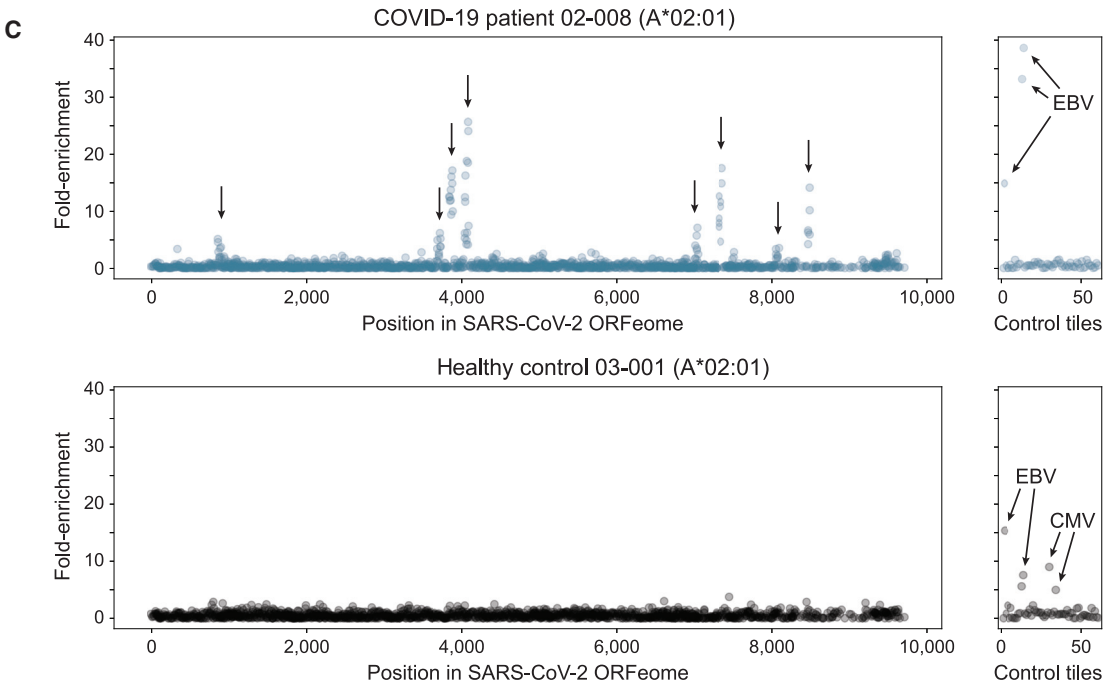
INTRODUCTION

Coronavirus disease 2019 (COVID-19) is a global pandemic that has claimed over a million lives worldwide and has affected countless more. Developing effective vaccines and therapies requires understanding how the virus and the immune response affect disease pathology and how the adaptive immune system recognizes and ultimately clears the virus. To date, most efforts have focused on the B-cell-mediated antibody response to the virus. Notably, the vast majority of current vaccine development efforts are focused on eliciting neutralizing antibodies to the virus, most frequently by immunizing with the spike (S) protein of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) or even with just the receptor binding domain (RBD) of the S protein (Vabret et al., 2020). How cytotoxic CD8+ T cells recognize and clear infected cells is less understood. In individuals who recovered from the closely related coronavirus SARS-CoV, virus-specific memory CD8+ T cells persist for at least 6–11 years, whereas memory B cells and anti-viral antibodies are largely undetectable at these later time points (Peng et al., 2006; Tang et al., 2011). Similarly, antibody responses to SARS-CoV-2 can be detected in most COVID-19 patients 10–15 days after symptom onset, but responses decline to baseline in many patients within 3 months (Seow et al., 2020). These findings suggest that vaccines focused solely on eliciting neutralizing antibodies to the S protein might be insufficient to elicit long-term immunity to coronaviruses. In mice infected with SARS-CoV, virus-specific CD8+ T cells are sufficient to enhance survival and diminish clinical disease (Zhao et al., 2010), and immunization with a single immunodominant CD8+ T cell epitope confers protection against lethal viral infection (Channappanavar et al., 2014). These studies highlight the importance of understanding the natural CD8+ T cell response to SARS-CoV-2 as a route to designing more durable vaccines.

SARS-CoV-2-specific CD8+ T cells can be detected in convalescent patients (Braun et al., 2020; Grifoni et al., 2020; Le Bert et al., 2020; Peng et al., 2006; Sekine et al., 2020; Thieme et al., 2020) and in subjects participating in vaccine trials (Folegatti et al., 2020; Jackson et al., 2020; Sahin et al., 2020). However, these studies used complex pools of predicted epitopes, and it is not clear which specific epitopes are being recognized and, in the case of vaccine trials, whether the epitopes being







(legend on next page)

recognized are the ones driving the natural CD8$^+$ T cell response to viral infection. To circumvent potential bias introduced by epitope prediction algorithms, we built on an unbiased, genome-wide screening technology (Kula et al., 2019) to simultaneously screen all of the memory CD8$^+$ T cells in convalescent patients against every possible epitope in SARS-CoV-2. We focused on memory cells to identify epitopes that are functionally recognized during the course of SARS-CoV-2 infection and included patients with a range of symptoms to determine whether any obvious associations are observed between CD8$^+$ T cell response and disease severity. CD8$^+$ T cells in these patients responded to a few highly antigenic epitopes in SARS-CoV-2 that were shared among patients with the same human leukocyte antigen (HLA) type. These epitopes were largely unique to SARS-CoV-2, were invariant among viral isolates, were frequently targeted by multiple clonotypes within each patient, and did not occur in "common cold" coronaviruses. Only ~10% of the epitopes were found in the S protein, whereas ~50% were located in ORF1ab and the highest density of epitopes were located in the nucleocapsid (N) protein. These results provide the necessary tools to better understand the CD8$^+$ T cell response in COVID-19 and provide a path to designing and developing next-generation vaccines.

RESULTS

Development of a Genome-Wide Screen for CD8$^+$ T Cell Epitopes in SARS-CoV-2

To determine the global landscape of CD8$^+$ T cell recognition in an unbiased fashion, we built on a genome-wide screening technology called T-Scan (Kula et al., 2019) that enabled us to simultaneously screen all memory CD8$^+$ T cells in a patient, one HLA allele at a time, against every possible viral epitope in SARS-CoV-2 as well as SARS-CoV and the four seasonal coronaviruses that cause the common cold (Figure 1A). Briefly, CD8$^+$ T cells were co-cultured with a genome-wide library of target cells (modified HEK293 cells), engineered to express a single HLA allele. Each target cell in the library also expressed a unique coronavirus-derived 61-amino acid (aa) protein fragment. These fragments were processed naturally by the target cells, and the appropriate peptide epitopes were displayed on major histocompatibility complex (MHC) class I molecules on the cell surface. When a CD8$^+$ T cell encountered its target in the co-culture, it secreted cytotoxic granules into the target cell, inducing apoptosis. Early apoptotic cells were then isolated from the co-culture, and the expression cassettes were sequenced, revealing the identity of the protein fragment. Because the assay is non-competitive, hundreds to thousands of T cells were screened simultaneously against tens of thousands of targets. To address the bottleneck of extensive sorting needed to isolate rare recognized target cells in high-complexity libraries (Kula et al., 2019), we engineered the target cells to express a Granzyme B (GzB)-activated fluorescent reporter as described previously as well as a GzB-activated version of the scramblase enzyme XKR8, which drives rapid and efficient transfer of phosphatidylserine to the outer membrane of early apoptotic cells (STAR Methods; Figure 1A). Early apoptotic cells were then enriched by magnetic-activated cell sorting with Annexin V, followed by fluorescence-activated sorting with the fluorescent reporter. This modification increased the throughput of the T-Scan assay 20-fold, enabling rapid processing of a large number of patient samples.

To comprehensively map responses to SARS-CoV-2, we generated a library of 61-aa protein fragments that tiled across all 11 open reading frames (ORFs) of SARS-CoV-2 in 20-aa steps (Figure 1B). To capture the known genetic diversity of SARS-CoV-2, we included all protein-coding variants from the 104 isolates that had been reported as of March 15, 2020. We also included the complete set of ORFs (ORFeome) of SARS-CoV and the four endemic coronaviruses that cause the common cold (betacoronaviruses HKU1 and OC43 and alphacoronaviruses NL63 and 229E). As positive controls, we included known immunodominant antigens from cytomegalovirus (CMV), Epstein-Barr virus (EBV), and influenza virus (Currier et al., 2002). Finally, each protein fragment was represented 10 times, each encoded with a unique nucleotide barcode to provide internal replicates in our screens, for a final library size of 43,420 clones.

Collection and Screening of Memory CD8$^+$ T Cells from COVID-19 Convalescent Patients

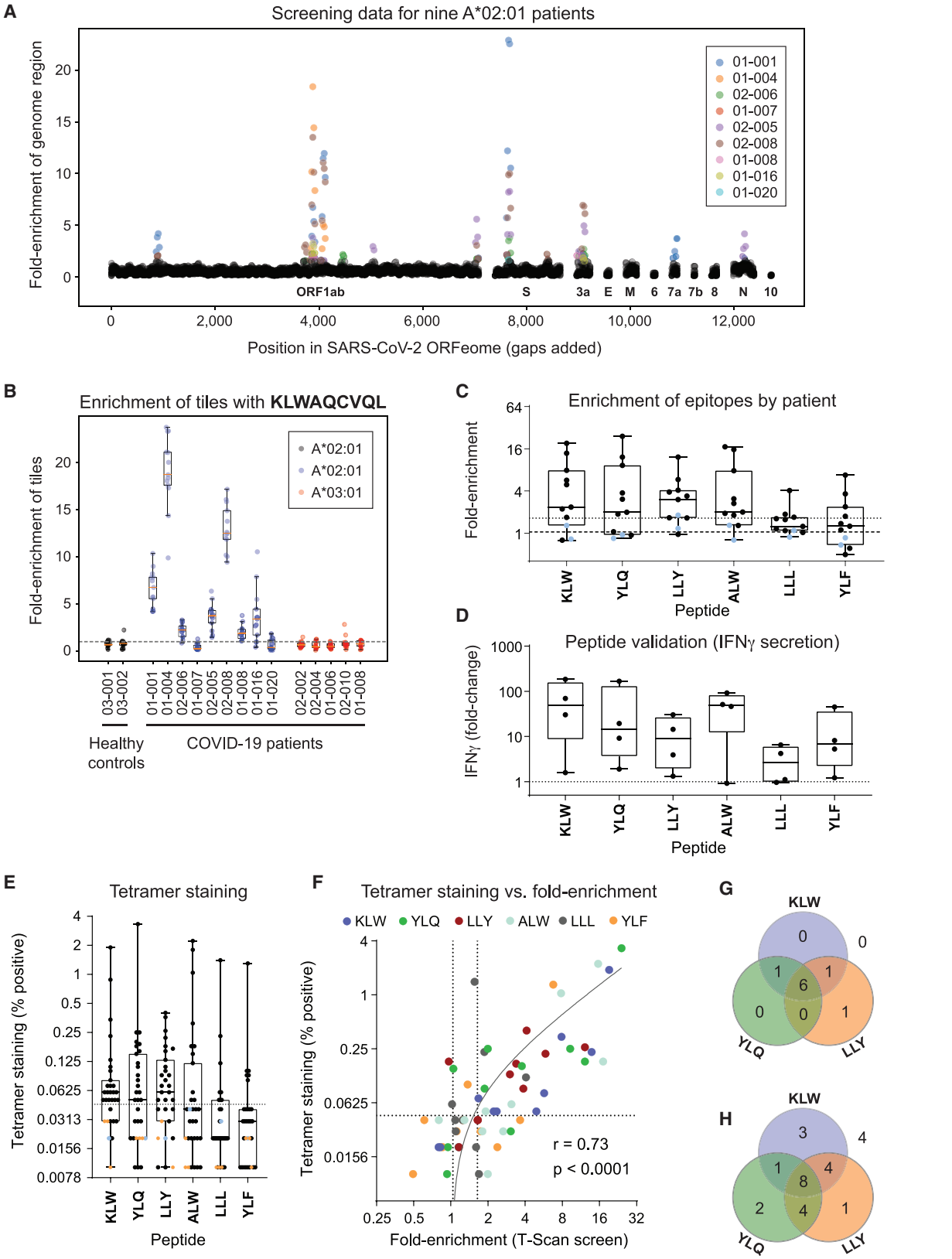
To identify the epitopes functionally recognized during the course of SARS-CoV-2 infection, we collected peripheral blood mononuclear cells from 78 adult patients who had tested positive by viral PCR (swab test), had recovered from their disease, and had been out of quarantine for at least 2 weeks according to Centers for Disease Control and Prevention guidelines (STAR Methods). Patients were recruited at either of two centers: Atlantic Heath System in Morristown, New Jersey and Ochsner Medical Center in New Orleans, Louisiana. All patients were HLA typed, and a summary of their characteristics and HLA types are provided in Tables S1 and S2. Because HLA A*02:01 is the most common MHC allele worldwide (Gonzalez-Galarza et al., 2020; Maiers et al., 2007), we started by selecting nine HLA-A*02:01 patients with a broad range of clinical presentations: six had mild symptoms and were not hospitalized, two required supplemental oxygen, and one required invasive ventilation. In each case, we purified bulk memory CD8$^+$ T cells (CD8$^+$, CD45RO$^+$, CD45RA$^-$, and CD57$^-$) by negative selection, expanded the cells with antigen-independent stimulation (anti-CD3), and screened them against the SARS-CoV-2 library.

Figure 1. T-Scan Approach for Comprehensive Mapping of the Memory CD8$^+$ T Cell Response to SARS-CoV-2

(A) Overview of the T-Scan antigen discovery screen.

(B) Design of the ORFeome-wide coronavirus antigen library.

(C) Example SARS-CoV-2 ORFeome-wide T-Scan screen data for a convalescent COVID-19 patient (top panel) and a healthy control individual (bottom panel). Each circle represents a single 61-aa SARS-CoV-2 protein fragment, with the x axis showing the position of each fragment in the concatenated SARS-CoV-2 ORFeome. The y axis shows the performance of the fragment in the screen, calculated as the ratio of sorted target cells expressing the protein fragment in relation to the unsorted target library. The right panels show the performance of the 60 positive control protein fragments derived from CMV, EBV, and influenza.



(legend on next page)

Target cells expressing only HLA-A*02:01 were used to provide unambiguous MHC restriction of discovered antigens. The SARS-CoV-2 screening results for one representative patient and one COVID-19$^-$ healthy control are shown in Figure 1C. We found reactivity to at least eight regions of SARS-CoV-2 proteins in the convalescent patient and none in the control. Importantly, we observed reproducible performance of four technical screen replicates, internal nucleic acid barcodes, and overlapping protein fragments, collectively suggesting robust screen performance. Additionally, we detected reactivity to the control CMV epitope (NLVPMVATV) in the healthy control, who was known to be CMV$^+$, and reactivity to two EBV epitopes in the COVID-19 patient and the healthy control individual (Figure 1C).

Identification and Validation of Shared CD8$^+$ T Cell Epitopes in SARS-CoV-2

Next, we examined the screen results for the full set of nine HLA-A*02:01 patients and detected reactivity to specific segments of SARS-CoV-2 ORFs in eight of nine patients (Figure 2A). In keeping with what has been observed for other viruses (Yewdell, 2006), we found that specific fragments of SARS-CoV-2 were recurrently recognized by the T cells of multiple patients (i.e., were immunodominant). For example, ORF1ab aa 3,881–3,900 and S aa 261–280 were each recognized by 7 of 9 patients (Figure 2A). Overall, we identified six regions that were targeted by CD8$^+$ T cells from at least three different patients. In addition to being shared across patients, these regions were among the strongest responses observed in each patient.

We next sought to identify the precise peptide epitopes underlying the shared T cell reactivities detected in our screens. The overlapping design of our antigen library allowed us to map the T cell reactivities to specific 20-aa segments. We then used NetMHC4.0 (Andreatta and Nielsen, 2016; Nielsen et al., 2003) to identify specific, high-affinity HLA-A*02:01 peptides in each pre-identified 20-aa stretch. An example of a predicted epitope and the corresponding screen data are shown in Figure 2B. Notably, the fragments scoring in our screens were enriched for high-affinity HLA-binding peptides compared with the library as a whole (Figure S1). To visualize the results across all nine patients, we collapsed the screening data into a single value (mean of screen replicates and redundant tiles), revealing a set of six predicted epitopes that were recurrently recognized by three or more patients (Figure 2C; Table 1). We then synthesized peptides corresponding to each epitope to validate our findings. All six epitopes induced peptide-dependent T cell activation as determined by interferon-γ (IFN-γ) secretion (Figure 2D) and CD137 upregulation (Figure S2). IFN-γ secretion and CD137 upregulation correlated with the fold enrichment in the T-Scan screens (Figures S2 and S3). As further validation, we constructed MHC tetramers with the 6 peptides and used them to stain the memory CD8$^+$ T cells of all 9 A*02:01 patients as well as an additional test set of 18 A*02:01 patients that had not been screened. Positive tetramer staining was observed in a subset of patients for all six peptides, including patients in the independent test set (Figure 2E). Additionally, the magnitude of enrichment in the screens correlated well with the frequency of cognate T cells in the patient samples ($r = 0.73$, $p < 0.0001$) (Figure 2F), allowing us to determine that our screens detected the targets of T cells that were present at 0.1% frequency or higher in the memory CD8$^+$ T cell pool. Notably, the 3 most commonly recognized epitopes we discovered—KLW, YLQ, and LLY—were each recognized by 67% of the patients we screened, and all 9 patients had a detectable response to at least 1 of the top 3 epitopes (Figure 2G). A similar analysis of the tetramer staining data in all 27 A*02:01 patients showed recognition of at least 1 of these epitopes in 23 of 27 patients (85%; Figure 2H). These analyses revealed the limited set of A*02:01-restricted shared epitopes recognized by patient T cells.

Because CD8$^+$ T cell responses are profoundly shaped by host MHC alleles, we next mapped memory CD8$^+$ T cell reactivities for five additional MHC alleles: HLA-A*01:01, HLA-A*03:01, HLA-A*11:01, HLA-A*24:02, and HLA-B*07:02. Along with

Figure 2. Discovery and Validation of Shared SARS-CoV-2 Epitopes Presented on HLA-A*02:01

(A) T-Scan screen data for nine HLA-A*02:01 COVID-19 patients. Each circle corresponds to a 20-aa segment of the SARS-CoV-2 ORFeome, with the x axis indicating the position of the segment in the ORFeome (gaps added for display purposes). The y axis shows the mean performance of all library fragments spanning the given 20-aa segment, calculated as described in Figure 1C. Results for each patient are denoted with different colors.

(B) Screen data for the KLW epitope (KLWAQCVQL). The boxplots represent the screen enrichments of all fragments in the library that contain the KLW epitope. Data for the nine HLA-A*02:01 COVID-19 patient screens are shown in blue, two healthy control individual HLA-A*02:01 screens are shown in gray, and five HLA-A*03:01 COVID-19 patient screens are shown in red.

(C) Collapsed screen data for six identified shared epitopes. Each boxplot shows the aggregate enrichment of one epitope in each of the nine screened HLA-A*02:01 patients (black dots) and two healthy control individuals (blue dots). The y axis shows the mean enrichment of all fragments in the library containing the given epitope. The dashed line indicates the mean enrichment, and a dotted line indicates the mean enrichment + 2 SD of all A*02:01 epitopes in all negative controls.

(D) IFN-γ ELISA validation of identified epitopes. HLA-A*02:01 target cells were pulsed with 1 μM peptide and incubated with memory CD8$^+$ T cells from four HLA-A*02:01 COVID-19 patients. The y axis shows the concentration of IFN-γ secreted by T cells from each patient (black dot) in the presence of each peptide compared with the no-peptide control. Data are the means of two technical replicates and representative of two independent experiments.

(E) Tetramer staining of memory CD8$^+$ T cells reactive to six shared HLA-A*02:01 epitopes. Memory CD8$^+$ T cells from 27 HLA-A*02:01 COVID-19 patients (black dots), one healthy HLA-matched control individual (blue dot), and three MHC-mismatched control individuals (orange dots) were stained with tetramers loaded with each of the six identified epitopes. The y axis indicates the percentage of tetramer-positive memory CD8$^+$ cells.

(F) Correlation of T-Scan screen performance and cognate T cell frequency as determined by tetramer staining. Each circle indicates the performance of one epitope in one of the nine screened HLA-A*02:01 patients. The x axis shows the aggregate performance of the epitope in the T-Scan screen, calculated as the average enrichment of all fragments containing that epitope. The y axis shows the frequency of tetramer-positive memory CD8$^+$ T cells recognizing each epitope.

(G and H) Recognition of the three most common HLA-A*02:01 epitopes across COVID-19 patients on the basis of (G) screening data (n = 9) or (H) tetramer staining (n = 27). For (G), patients were considered positive for an epitope when the aggregate performance of the epitope in the screen data exceeded a set threshold (mean + 2SD of the enrichment of all of the SARS-CoV-2 fragments in the healthy control individuals). For (H), patients were considered positive for an epitope when 0.05% or more of memory CD8$^+$ T cells were positive by tetramer staining.

Table 1. Shared CD8$^+$ T Cell Epitopes Identified in COVID-19 Convalescent Patients

	Allele	Peptide Name	Full Peptide	Parent Protein	Start	End	Affinity[a] (nM)	% of Patients (Screen)
1	A*02:01	KLW	KLWAQCVQL	ORF1ab	3,886	3,894	17.7	88.9
2	A*02:01	YLQ	YLQPRTFLL	S	269	277	5.4	77.8
3	A*02:01	LLY	LLYDANYFL	ORF3a	139	147	3.1	88.9
4	A*02:01	ALW	ALWEIQQVV	ORF1ab	4,094	4,102	7.8	88.9
5	A*02:01	LLL	LLLDRLNQL	N	222	230	14.8	33.3
6	A*02:01	YLF	YLFDESGEFKL	ORF1ab	906	916	22.2	44.4
7	A*01:01	FTS	FTSDYYQLY	ORF3a	207	215	3.2	100
8	A*01:01	TTD	TTDPSFLGRY	ORF1ab	1,637	1,646	7.2	100
9	A*01:01	PTD	PTDNYITTY	ORF1ab	1,321	1,329	6.1	80
10	A*01:01	ATS	ATSRTLSYY	M	171	179	16.7	60
11	A*01:01	CTD	CTDDNALAYY	ORF1ab	4,163	4,172	5.3	100
12	A*01:01	NTC	NTCDGTTFTY	ORF1ab	4,082	4,091	121.8	60
13	A*01:01	DTD	DTDFVNEFY	ORF1ab	5,130	5,138	2.8	40
14	A*01:01	GTD	GTDLEGNFY	ORF1ab	3,437	3,445	6	40
15	A*03:01	KTF	KTFPPTEPK	N	361	369	20.8	100
16	A*03:01	KCY	KCYGVSPTK	S	378	386	152.6	100
17	A*03:01	VTN	VTNNTFTLK	ORF1ab	808	816	19.8	40
18	A*03:01	KTI	KTIQPRVEK	ORF1ab	282	290	113.2	40
19	A*11:01	KTF	KTFPPTEPK	N	361	369	6.3	100
20	A*11:01	VTD	VTDTPKGPK	ORF1ab	4,216	4,224	160.6	60
21	A*11:01	ATE	ATEGALNTPK	N	134	143	55.5	80
22	A*11:01	ASA	ASAFFGMSR	N	311	319	14.4	40
23	A*11:01	ATS	ATSRTLSYYK	M	171	180	7.9	60
24	A*24:02	QYI	QYIKWPWYI	S	1,208	1,216	13.2	60
25	A*24:02	VYF	VYFLQSINF	ORF3a	112	120	47.4	80
26	A*24:02	VYI	VYIGDPAQL	ORF1ab	5,721	5,729	206	40
27	B*07:02	SPR	SPRWYFYYL	N	105	113	6.3	80
28	B*07:02	RPD	RPDTRYVL	ORF1ab	2,949	2,956	56.9	80
29	B*07:02	IPR	IPRRNVATL	ORF1ab	5,916	5,924	5.1	20

[a]Affinity (equilibrium dissociation constant) predicted by using NetMHC4.0.

HLA-A*02:01, these alleles provide a broad perspective regarding the nature of anti-SARS-CoV-2 CD8$^+$ T cell immunity because ~90% of the United States population and ~85% of the world population are positive for at least 1 of the 6 alleles we examined (Gonzalez-Galarza et al., 2020; Maiers et al., 2007). For each allele, we selected five HLA$^+$ convalescent COVID-19 patients and screened their memory CD8$^+$ T cells against the SARS-CoV-2 library in target cells expressing only the single HLA of interest. Because some patients were positive for more than one allele, their T cells were used in more than one HLA-specific screen. A total of 25 distinct patients was needed for the 34 HLA-specific screens. As with A*02:01 patients, we found robust T cell recognition of multiple regions in the SARS-CoV-2 ORFeome for patients with each HLA allele (Figure S4) and confirmed that the scoring fragments were enriched for predicted high-affinity MHC binders for each respective allele (Figure S1). We again observed recurrent recognition of specific protein fragments by most or all patients for each allele (Figure 3A), indicating a narrow set of shared responses. As before, we used NetMHC4.0 to identify the precise epitopes underlying the top hits from our screens, and validated these peptides by using IFN-γ secretion (Figure 3B) and CD137 upregulation (Figure S2). We identified 3 or more recurrently recognized epitopes for each screened MHC allele and found that 92% of patients recognized at least 1 of the top 3 allele-specific epitopes for these 5 additional HLA types (Figure 3C). Collectively, we mapped and validated 29 CD8$^+$ T cell epitopes that were shared among COVID-19 patients with the same HLA type (Table 1). These epitopes represent the global landscape of MHC class I immunodominance in SARS-CoV-2 across the six most prevalent HLA types.

Most Shared CD8$^+$ T Cell Epitopes Reside Outside of the S Protein

The unbiased antigen mapping we performed enabled us to interrogate various features of CD8$^+$ T cell immunity to SARS-CoV-2. First, we examined the scope of recognized viral proteins. We observed broad reactivity to many SARS-CoV-2 proteins,

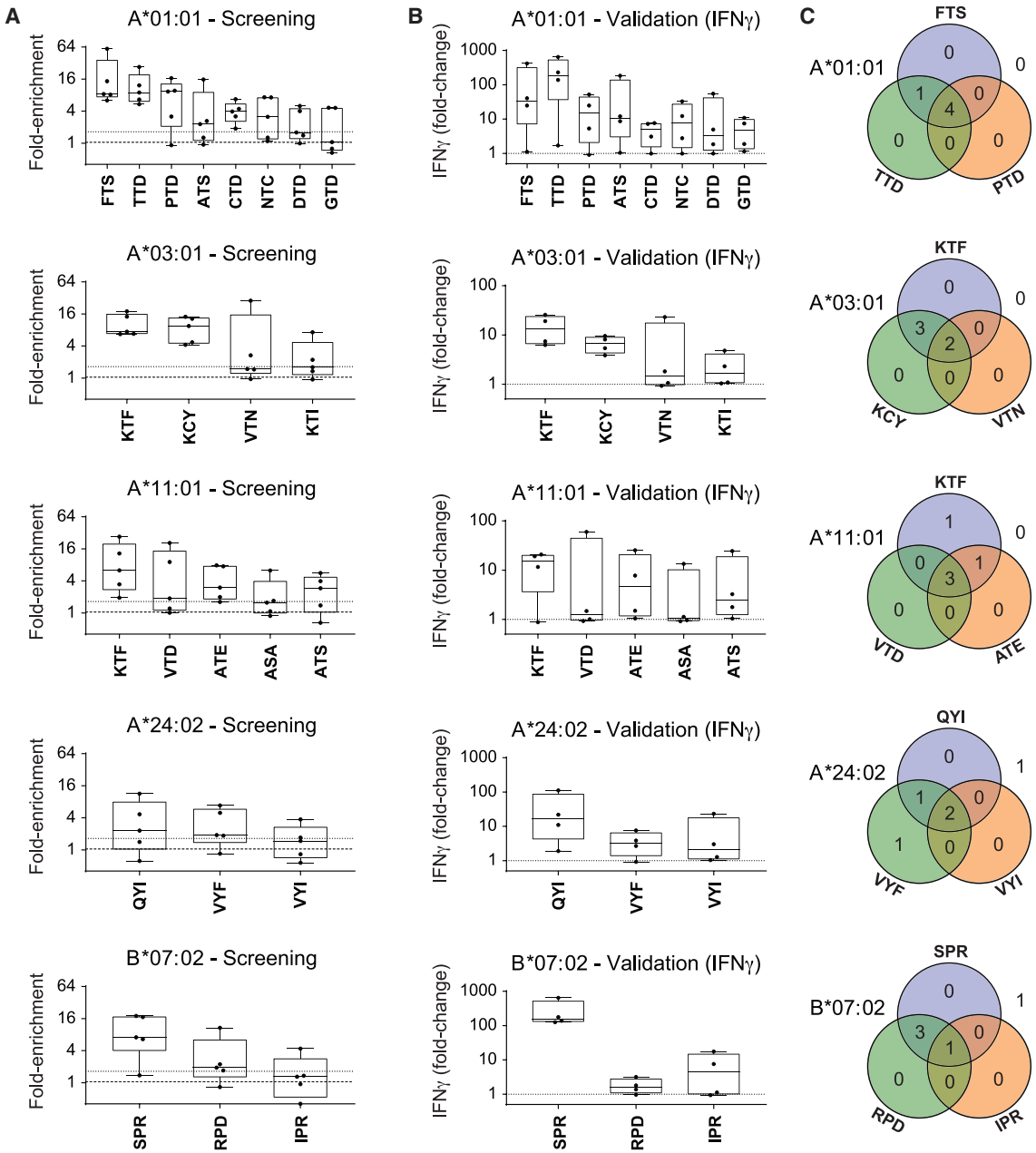


Figure 3. Discovery and Validation of Shared SARS-CoV-2 Epitopes Presented on HLA-A*01:01, HLA-A*03:01, HLA-A*11:01, HLA-A*24:02, and HLA-B*07:02

(A) Collapsed T-Scan screen data for shared epitopes identified for each analyzed MHC allele. Each boxplot shows the aggregate enrichment of one epitope in each of the five COVID-19 patients (black dots) screened for the listed allele. The y axis shows the mean enrichment of all fragments in the library containing the given epitope. Full epitope sequences are listed in Table 1.

(B) IFN-γ ELISA validation of identified epitopes. Memory $CD8^+$ T cells from four COVID-19 patients positive for each prioritized MHC allele were incubated with MHC-matched target cells pulsed with 1 μM peptide. The y axis shows the concentration of IFN-γ secreted by T cells from each patient (black dot) in the presence of each peptide compared with a no-peptide control. Data are the means of two technical replicates and representative of two independent experiments.

(C) Recognition of the three most common epitopes for each prioritized MHC allele across five COVID-19 patients. Patients were considered positive for an epitope when the aggregate performance of the epitope in the screen data exceeded a set threshold (mean + 2SD of the enrichment of all of the SARS-CoV-2 fragments in the healthy controls).

including ORF1ab, S, N, M, and ORF3a (Figure 4A). Notably, only 3 of the 29 epitopes were located in the S protein. Most epitopes (15 of 29) were located in ORF1ab, and the highest density of epitopes were located in the N protein. When taken in aggregate, our results are consistent with previous ORF-level analyses using peptide pools (Altmann and Boyton, 2020; Braun et al., 2020;

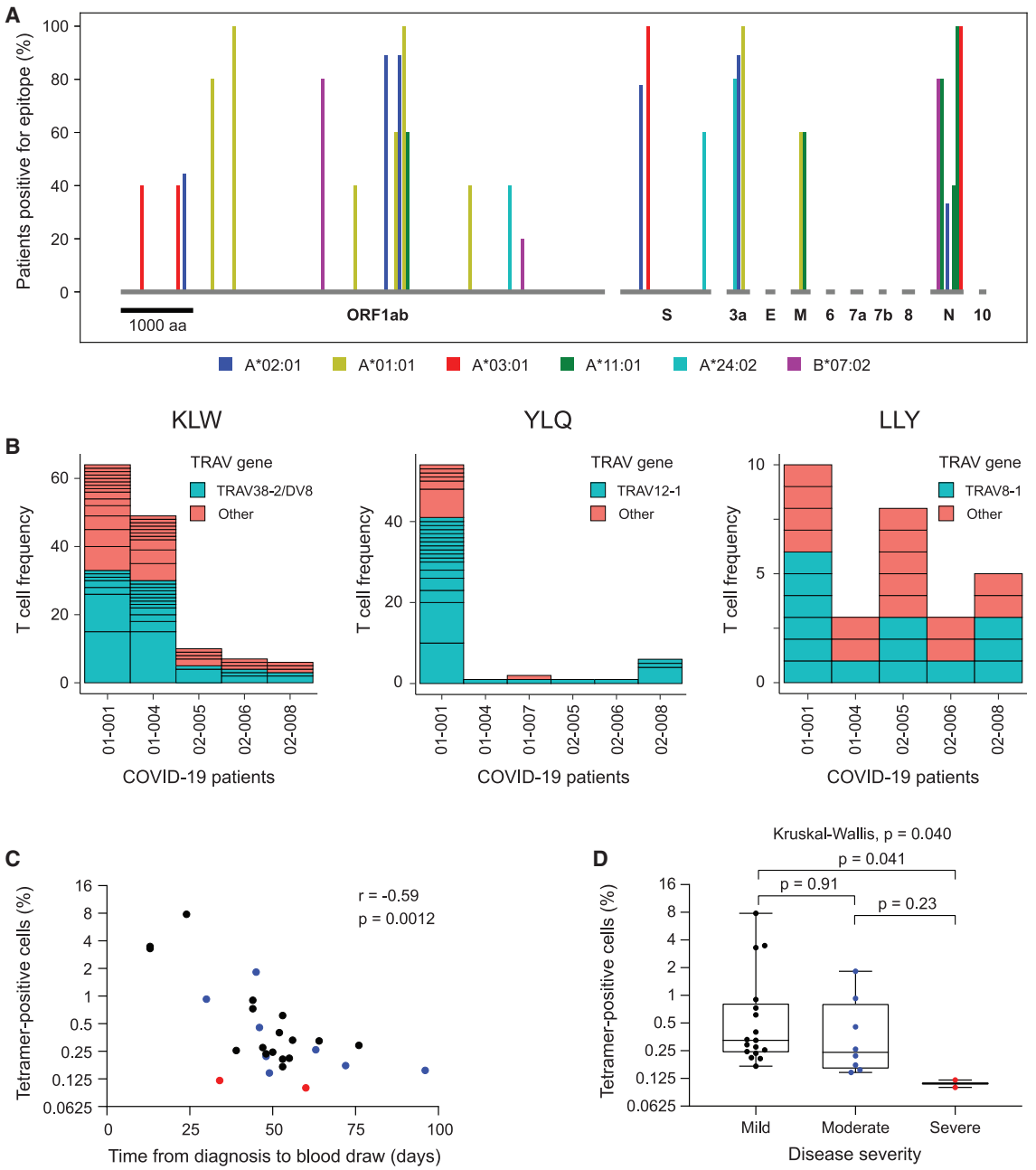


Figure 4. Shared Epitopes Span the SARS-CoV-2 ORFeome and Are Recognized by TCRs with Common Features

(A) Distribution of shared CD8$^+$ T cell epitopes across the SARS-CoV-2 genome. Each bar represents one validated shared epitope, with the x axis showing its position in the SARS-CoV-2 ORFeome, the color indicating its MHC restriction, and the height of the bar indicating the percentage of MHC-matched patients recognizing the epitope. Patients were considered positive for an epitope when the aggregate performance of the epitope in the screen data exceeded a threshold (mean + 2SD of the enrichment of all SARS-CoV-2 fragments in the healthy controls). For clarity, overlapping epitopes are plotted as adjacent bars.

(B) TCR alpha variable (TRAV) gene usage in tetramer$^+$ T cells across patients. The height of each box corresponds to the number of T cells within the clonotype. Blue corresponds to the conserved TRAV gene for a specific epitope, and red corresponds to all other TRAV genes.

(C) The magnitude of detected memory CD8$^+$ T cell response correlates negatively with time from diagnosis to blood draw. The p value and correlation coefficient r were calculated using Pearson correlation.

(D) Magnitude of detected memory T cell response in patients with mild, moderate, and severe disease.

For (C) and (D), the magnitude of T cell response indicates the total fraction of memory CD8$^+$ T cells that stained positive with tetramers for one of the six identified HLA-A*02:01 epitopes (KLW, YLQ, LLY, ALW, LLL, and YLF) for 27 HLA-A*02:01 patients.

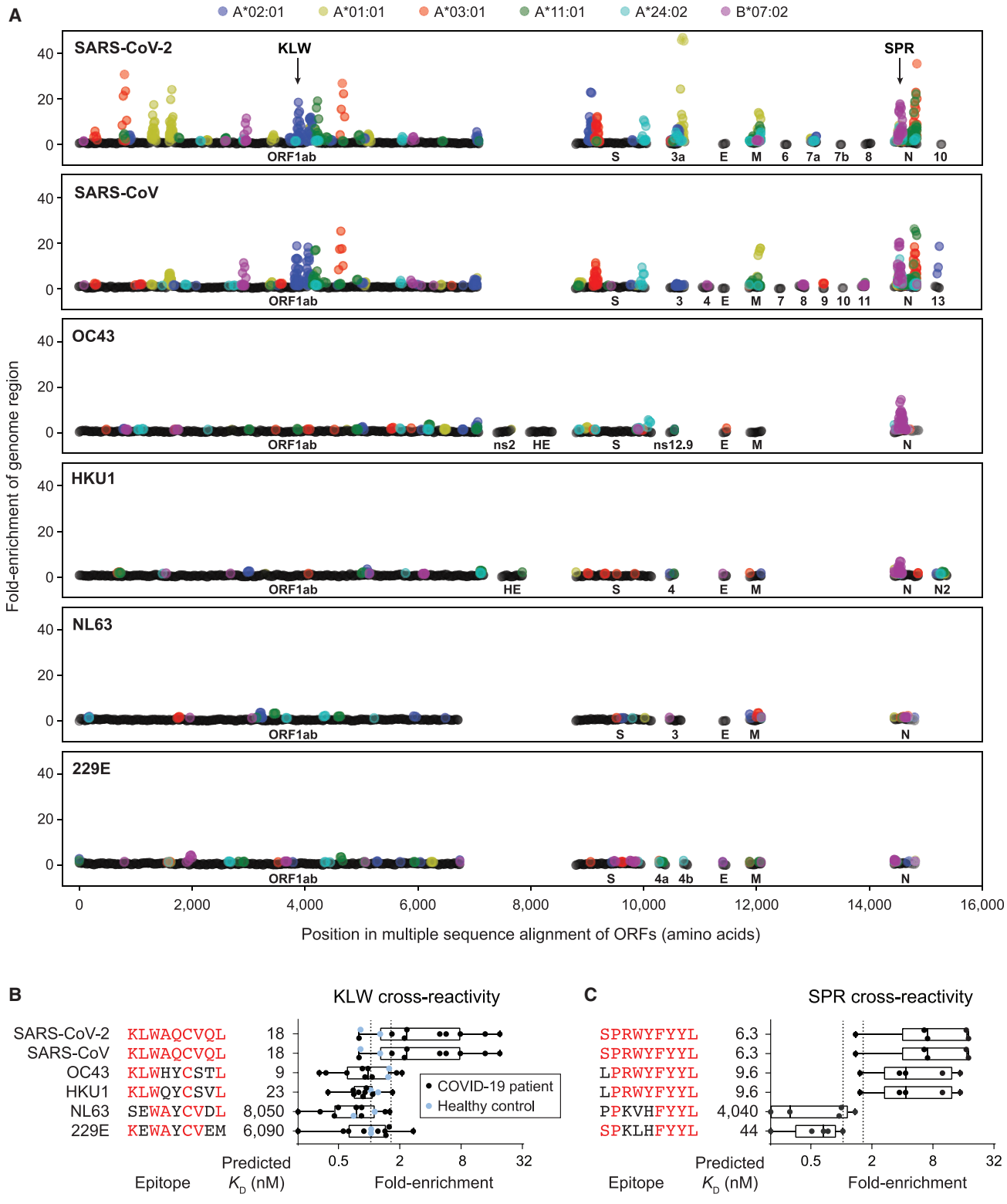


Figure 5. Minimal Cross-reactivity of SARS-CoV-2-Reactive Memory CD8⁺ T Cells with Other Coronaviruses

(A) Screen data compared across coronavirus ORFeomes. Each graphic shows the collective reactivity to one coronavirus genome (SARS-CoV-2, SARS-CoV-1, OC43, HKU1, NL63, or 229E) detected in the 34 T-Scan screens performed. Each circle corresponds to a 20-aa stretch of the coronavirus ORFeome, with the x axis indicating the position of the stretch in the ORFeome. The y axis shows the mean performance of all of the library fragments spanning the given 20-aa stretch, calculated as described in Figure 1C. Results include nine HLA-A*02:01 patients, five HLA-A*03:01 patients, five HLA-A*01:01 patients, five HLA-A*11:01

(legend continued on next page)

Grifoni et al., 2020; Le Bert et al., 2020; Thieme et al., 2020). However, our approach provided an increased level of granularity that enabled identification of specific epitope sequences and highlighted HLA-allele-specific differences. For example, we observed shared epitopes in the S protein for HLA-A*02:01, HLA-A*03:01, and HLA-A*24:02 but not for HLA-A*01:01, HLA-A*11:01, or HLA-B*07:02. Notably, we detected only one recurrent response in the RBD of the S protein (KCY on HLA-A*03:01).

Next, we wanted to determine how the CD8$^+$ T cell response to SARS-CoV-2 intersects with the emerging genetic diversity of the virus. Recent analyses that examined the genome sequences of over 10,000 isolates of SARS-CoV-2 sampled from 68 different countries identified a set of 28 non-synonymous coding mutations detected in at least 1% of strains (Koyama et al., 2020). Only one of these mutations (M protein T175M, detected in 2% of strains) was found in the shared epitopes we identified (HLA-A*01:01 ATS and HLA-A*11:01 ATS). This suggests that recognition of the epitopes we identified is unlikely to be influenced significantly by the SARS-CoV-2 genetic diversity observed so far.

Structural Basis of Epitope Recognition in SARS-CoV-2

Identifying specific SARS-CoV-2 epitopes allowed us to examine the features of the T cell receptors (TCRs) recognizing these shared epitopes. We used tetramers loaded with three HLA-A*02:01 epitopes (KLW, YLQ, and LLY) to stain and sort antigen-specific memory CD8$^+$ T cells from the initial nine HLA-A*02:01$^+$ convalescent COVID-19 patients. For each of the other 5 HLA alleles, we used tetramer or CD137 staining to sort CD8$^+$ T cells reactive to the 3–4 most frequently shared epitopes in 2 patients each. We then used 10x Genomics single-cell sequencing to identify the paired TCR α and TCR β chains expressed by these T cells. Collectively, we found TCRs recognizing 17 shared epitopes for a total of 421 SARS-CoV-2-reactive TCRs. Next, we examined the TCR sequences themselves, focusing on the three HLA-A*02:01 antigens that were explored across a larger set of patients. We identified paired clonotypes reactive to each antigen in 5 of 9 (KLW and ALW) or 6 of 9 (YLQ) patients. For the majority of responses (9 of 16), we detected oligoclonal recognition by 5 or more distinct clonotypes. Striking similarity was observed among the TCRs recognizing each antigen in terms of Vα gene segment usage and, to a lesser extent, Vβ usage (Figure 4B). Specifically, 26 of 61 KLW-reactive clonotypes used TRAV38-2/DV8, 24 of 31 YLQ-reactive clonotypes used TRAV12-1, and 14 of 29 LLY-reactive clonotypes used TRAV8-1. Notably, these dominant Vα genes were used across all of the patients for whom we identified reactive clonotypes. Altogether, these data suggest that certain TCR Vα regions provide the structural features necessary for high-affinity binding to peptide-MHC and that these features might explain the recurrent recognition of these epitopes among patients with the same HLA type.

Correlations of SARS-CoV-2-Specific T Cells with Clinical Characteristics

Although our study was not designed to test specific clinical hypotheses, we looked for potential associations between virus-specific T cell responses and clinical characteristics. We focused on the 27 A*02:01 patients for whom tetramer staining data were available because this represents the most uniform set of T cell data in our study. No obvious association was observed between T cell response and sex (Figure S5), but a significant negative correlation was observed with time from diagnosis to blood draw ($p = 0.0012$) (Figures 4C and S5). This is expected because anti-viral T cells, including effector memory cells, naturally contract after acute infection (Badovinac et al., 2002; Wherry and Ahmed, 2004). This observation is important, however, because future epidemiological studies should control for this variable. We also observed a trend where patients with severe disease exhibited fewer virus-specific T cells than those with mild disease ($p = 0.041$) (Figures 4D and S5). Additionally, older patients had lower T cell responses than younger patients (Figure S5). These observations should be treated with caution because the number of patients in these studies is small, particularly those requiring invasive ventilation. Appropriately powered studies to address this question are warranted because they could shed light on whether these shared epitopes are potentially protective against severe disease.

COVID-19 Memory CD8$^+$ T Cells Show Minimal Cross-Reactivity with Endemic Coronaviruses

Another key question is how pre-existing immunity to other coronaviruses shapes the CD8$^+$ T cell response to SARS-CoV-2. There are four commonly circulating coronaviruses (OC43, HKU1, NL63, and 229E), and cross-reactive responses to these viruses have been theorized to be a potential protective factor during SARS-CoV-2 infection (Sette and Crotty, 2020). Moreover, understanding the extent of cross-reactivity has implications for accurately monitoring T cell responses to SARS-CoV-2 and for optimizing vaccine design. If the immune response to SARS-CoV-2 is shaped by pre-existing CD8$^+$ T cells that recognize other coronaviruses, we reasoned that COVID-19 patients should have reactivity to regions of the other coronaviruses that correspond to the SARS-CoV-2 epitopes we identified. We therefore examined T cell reactivity to SARS-CoV-2, SARS-CoV, and all 4 endemic coronaviruses in the 34 genome-wide screens that we conducted across all patients and MHC alleles (Figure 5A). We observed broad reactivity to the corresponding epitopes in SARS-CoV in over half of cases, consistent with a recent study reporting the existence of long-lasting memory T cells cross-reactive to SARS-CoV-2 in patients who had been

patients, five HLA-A*24:02 patients, and five HLA-B*07:02 patients. For visualization, the positions of the conserved ORF1ab, S, M, E, and N proteins were aligned across all ORFeomes.

(B) Alignment of the KLW epitope across coronavirus genomes. The alignment shows the region of each coronavirus genome corresponding to the SARS-CoV-2 HLA-A*02:01 KLW epitope. The boxplots show the aggregate screen performance of all fragments containing each epitope variant for nine HLA-A*02:01-positive COVID-19 patients (black dots) and two HLA-A*02:01-positive healthy controls (blue dots).

(C) Alignment of the SPR epitope across coronavirus genomes. The alignment shows the region of each coronavirus genome corresponding to the SARS-CoV-2 HLA-B*07:02 epitope. The boxplots show the aggregate screen performance of all fragments containing each epitope variant for five HLA-B*07:02-positive COVID-19 patients (black dots).

infected with SARS-CoV during the 2002/2003 SARS outbreak (Le Bert et al., 2020). In contrast, we detected almost no reactivity to OC43 and HKU1 (2 of 29 dominant epitopes) and none to NL63 and 229E. Beyond the 29 epitopes, we observed no reproducible cross-reactivity to any other regions of the four endemic coronaviruses, again suggesting that prior exposure to these viruses is unlikely to provide $CD8^+$ T-cell-based protection from SARS-CoV-2.

Mapping the specific shared epitopes in SARS-CoV-2 enabled us to determine the molecular basis for this lack of cross-reactivity. In some cases, the corresponding region is poorly conserved in the other coronaviruses, and high-affinity binding to MHC is lost (see, for example, the corresponding regions of the KLW epitope in NL63 and 229E; Figure 5B). In other cases, the corresponding epitopes are still predicted to bind with high affinity to MHC, but SARS-CoV-2-reactive T cells did not recognize them (see, for example, the corresponding regions of the KLW epitope in OC43 and HKU1) (Figure 5B). In one notable case, we did identify a strong cross-reactive response. The HLA B*07:02 epitope SPR, which lies in the N protein, is highly conserved across betacoronaviruses, and all four of the patients who demonstrated reactivity to SPR also exhibited reactivity to the corresponding epitopes in OC43 and HKU1 (Figure 5C). Overall, however, we conclude that the $CD8^+$ T cell response to SARS-CoV-2 is not significantly shaped by pre-existing immunity to endemic coronaviruses.

DISCUSSION

Although peptide pools can reveal $CD8^+$ T cell responses in COVID-19 patients (Altmann and Boyton, 2020; Braun et al., 2020; Grifoni et al., 2020; Le Bert et al., 2020; Thieme et al., 2020), they do not identify which specific epitopes are being recognized. Our use of unbiased, genome-wide screens revealed a relatively small set of shared epitopes recognized by the memory $CD8^+$ T cells of convalescent patients. These epitopes varied across HLA types in the total number of recognized epitopes and in the proteins in which they reside. This emphasizes the importance of searching for MHC associations with disease outcome in COVID-19 patients and of detailed tracking of MHC alleles in immune monitoring of vaccine trials.

We estimate that our screens detected T cell specificities that were present at 0.1% frequency or higher in the pool of memory $CD8^+$ T cells for each patient. There are likely other virus-specific T cells below this threshold, and these T cells can be studied in future experiments by antigen-specific expansion by using pools of peptides. However, the epitopes detected in this study represented the targets of the most expanded clones and so are likely to have played the largest role in the patient's $CD8^+$ T cell response to SARS-CoV-2 infection. Knowing the identity of these epitopes provides a way to rapidly monitor SARS-CoV-2-reactive T cells in patients with various disease courses and to assess the effectiveness of vaccines in eliciting responses that mimic natural infection. Additionally, measuring reactivity to the epitopes we identified can serve as the basis of a T-cell-based diagnostic, similar to the IFN-γ release assay used for tuberculosis testing (Albert-Vega et al., 2018). T-cell-based testing could be an important complement to serological testing because recent reports have shown a rapid decline in immunoglobulin G (IgG) responses after viral infection, with 40% of asymptomatic individuals becoming seronegative within 3 months (Long et al., 2020). Additionally, memory T cells might persist longer than antibodies or memory B cells, as observed for SARS-CoV (Le Bert et al., 2020; Ng et al., 2016; Peng et al., 2006; Tang et al., 2011).

In contrast to reports that unexposed individuals have T cells that cross-react with SARS-CoV-2 (Grifoni et al., 2020; Le Bert et al., 2020; Mateus et al., 2020; Weiskopf et al., 2020), our data showed that the most expanded memory T cells in convalescent patients did not cross-react with endemic coronaviruses. These discordant results might be explained by our focus on $CD8^+$ T cell responses, whereas most cross-reactivity is detected in $CD4^+$ T cells. Additionally, it is possible that weakly cross-reactive T cells exist in unexposed individuals but that these T cells are overtaken by *de novo* responses during SARS-CoV-2 infection. If pre-existing memory responses to other coronaviruses efficiently recognize SARS-CoV-2, then the reacting T cells should expand, and their targets would likely have been detected in our screens. As a result, the paucity of identified cross-reactive responses argues against substantial protection against SARS-CoV-2 stemming from $CD8^+$ T cell immunity to the four coronaviruses that cause the common cold. We did identify two epitopes that were shared with OC43 and HKU1, which could be of interest in the design of vaccines intended to boost pre-existing T cell immunity.

Our findings have broader implications for SARS-CoV-2 vaccine design. The vast majority of shared epitopes we uncovered (26 of 29) were located in ORF1ab, N, M, and ORF3a; only 3 were in S, and only 1 was in the RBD of S. These findings provide high-resolution insights into peptide pool studies observing responses outside of the S protein and are consistent with the detectable but modest $CD8^+$ T cell responses generated by vaccines targeting the S protein (Grifoni et al., 2020; Le Bert et al., 2020; Mulligan et al., 2020). Importantly, the protective or pathogenic role of $CD8^+$ T cell responses to specific proteins, individual shared epitopes, or epitopes that are only recognized after vaccination remains to be determined. The epitopes we identified can serve as the basis of experimental and correlational studies to address this critical question. Moreover, our findings enable the design and evaluation of next-generation vaccines that more fully recapitulate the scope of natural $CD8^+$ T cell responses to SARS-CoV-2 infection.

Limitations of Study

Although our screening approach assayed all patient memory $CD8^+$ T cells as a pool, it is best suited for discovery of targets recognized by the most abundant T cell specificities ($\geq$0.1% based on our estimates). Additional specificities recognized by less frequent T cell clonotypes could have been missed. In addition, sample limitations necessitated polyclonal expansion of the memory $CD8^+$ T cells *ex vivo* that might have altered the relative abundance of some clonotypes. Finally, our study was underpowered to evaluate the clinical effect of $CD8^+$ T cells recognizing specific epitopes. Additional studies are needed to determine whether $CD8^+$ T cell responses to individual proteins or epitopes are associated with protection from the virus or specific clinical outcomes.

Ethics Statement

All donors provided written consent. The study was conducted in accordance with the Declaration of Helsinki (1996), approved by the Atlantic Health System Institutional Review Board and the Ochsner Clinic Foundation Institutional Review Board, and registered at ClinicalTrials.gov (NCT04397900). Details regarding sample collection design and all other methods are provided in the STAR Methods.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Ethics statement
  - Sample collection design
  - Patient recruitment and demographics
- METHOD DETAILS
  - Isolation of PBMCs and Memory CD8+ T cells
  - Peptide Library Generation and Cloning
  - TScan Screen
  - Peptide Validation Assay
  - Tetramer Staining
  - Single Cell TCR sequencing
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.immuni.2020.10.006.

ACKNOWLEDGMENTS

We thank the patients and their families who participated in this study. We also thank Stephen Elledge and Kai Wucherpfennig for helpful discussions. This work was supported by TScan Therapeutics, a privately owned biotechnology company.

AUTHOR CONTRIBUTIONS

A.P.F., T.K., and G.M. conceived the project. A.P.F., T.K., Y.W., D.M.V.N., A.W., G.S.D., Q.X., N.N., C.R.P., H.J.W., A.V., and G.M. designed, performed, and analyzed experiments. T.K., A.F., A.W.C., Q.X., and G.S.D. designed and performed bioinformatics analyses. L.R.B., A.T.A., and E.D.W. directed sample collection and recruited patients. S.C. designed the sample collection study. K.J.O. and S.A.B. oversaw sample collection and tracking for TScan Therapeutics. A.P.F., T.K., S.A.B., S.C., and G.M. wrote the manuscript.

DECLARATION OF INTERESTS

A.P.F., Y.W., D.M.V.N., A.W., G.S.D., Q.X., N.N., C.R.P., A.W.C., H.J.W., A.V., K.J.O., S.A.B., and G.M. are employees of TScan Therapeutics. T.K. is a founder of TScan Therapeutics, and T.K. and G.M. hold equity in TScan Therapeutics. S.C. is a consultant for TScan Therapeutics. A patent application relating to the T-Scan technology used in the current manuscript was filed previously by The Brigham and Women's Hospital, Inc. Provisional patent applications covering SARS-CoV-2 epitope sequences and TCR sequences described in the current manuscript as well as related uses, diagnostics, therapeutics, and vaccines for COVID-19 were filed by TScan Therapeutics, Inc.

Received: July 31, 2020
Revised: September 3, 2020
Accepted: October 13, 2020
Published: October 20, 2020

REFERENCES

Albert-Vega, C., Tawfik, D.M., Trouillet-Assant, S., Vachot, L., Mallet, F., and Textoris, J. (2018). Immune Functional Assays, From Custom to Standardized Tests for Precision Medicine. Front. Immunol. *9*, 2367.

Altmann, D.M., and Boyton, R.J. (2020). SARS-CoV-2 T cell immunity: Specificity, function, durability, and role in protection. Sci. Immunol. *5*, eabd6160.

Andreatta, M., and Nielsen, M. (2016). Gapped sequence alignment using artificial neural networks: application to the MHC class I system. Bioinformatics *32*, 511–517.

Badovinac, V.P., Porter, B.B., and Harty, J.T. (2002). Programmed contraction of CD8(+) T cells after infection. Nat. Immunol. *3*, 619–626.

Braun, J., Loyal, L., Frentsch, M., Wendisch, D., Georg, P., Kurth, F., Hippenstiel, S., Dingeldey, M., Kruse, B., Fauchere, F., et al. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. Nature. Published online July 29, 2020. https://doi.org/10.1038/s41586-020-2598-9.

Channappanavar, R., Fett, C., Zhao, J., Meyerholz, D.K., and Perlman, S. (2014). Virus-specific memory CD8 T cells provide substantial protection from lethal severe acute respiratory syndrome coronavirus infection. J. Virol. *88*, 11034–11044.

Currier, J.R., Kuta, E.G., Turk, E., Earhart, L.B., Loomis-Price, L., Janetzki, S., Ferrari, G., Birx, D.L., and Cox, J.H. (2002). A panel of MHC class I restricted viral peptides for use as a quality control for vaccine trial ELISPOT assays. J. Immunol. Methods *260*, 157–172.

Folegatti, P.M., Ewer, K.J., Aley, P.K., Angus, B., Becker, S., Belij-Rammerstorfer, S., Bellamy, D., Bibi, S., Bittaye, M., Clutterbuck, E.A., et al.; Oxford COVID Vaccine Trial Group (2020). Safety and immunogenicity of the ChAdOx1 nCoV-19 vaccine against SARS-CoV-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. Lancet *396*, 467–478.

Gonzalez-Galarza, F.F., McCabe, A., Santos, E.J.M.D., Jones, J., Takeshita, L., Ortega-Rivera, N.D., Cid-Pavon, G.M.D., Ramsbottom, K., Ghattaoraya, G., Alfirevic, A., et al. (2020). Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. Nucleic Acids Res. *48* (D1), D783–D788.

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.

Jackson, L.A., Anderson, E.J., Rouphael, N.G., Roberts, P.C., Makhene, M., Coler, R.N., McCullough, M.P., Chappell, J.D., Denison, M.R., Stevens, L.J., et al.; mRNA-1273 Study Group (2020). An mRNA vaccine against SARS-CoV-2 - preliminary report. N. Engl. J. Med. Published online July 14, 2020. https://doi.org/10.1056/NEJMoa2022483.

Koyama, T., Platt, D., and Parida, L. (2020). Variant analysis of SARS-CoV-2 genomes. Bull. World Health Organ. *98*, 495–504.

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.e3.

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.

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.

Maiers, M., Gragert, L., and Klitz, W. (2007). High-resolution HLA alleles and haplotypes in the United States population. Hum. Immunol. *68*, 779–788.

Mateus, J., Grifoni, A., Tarke, A., Sidney, J., Ramirez, S.I., Dan, J.M., Burger, Z.C., Rawlings, S.A., Smith, D.M., Phillips, E., et al. (2020). Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. Science *370*, 89–94.

Mulligan, M.J., Lyke, K.E., Kitchin, N., Absalon, J., Gurtman, A., Lockhart, S.P., Neuzil, K., Raabe, V., Bailey, R., and Swanson, K.A. (2020). Phase 1/2 study to describe the safety and immunogenicity of a COVID-19 RNA vaccine candidate (BNT162b1) in adults 18 to 55 years of age: interim report. medRxiv. https://doi.org/10.1101/2020.06.30.20142570.

Ng, O.W., Chia, A., Tan, A.T., Jadi, R.S., Leong, H.N., Bertoletti, A., and Tan, Y.J. (2016). Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. Vaccine *34*, 2008–2014.

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.

Peng, H., Yang, L.T., Wang, L.Y., Li, J., Huang, J., Lu, Z.Q., Koup, R.A., Bailer, R.T., and Wu, C.Y. (2006). Long-lived memory T lymphocyte responses against SARS coronavirus nucleocapsid protein in SARS-recovered patients. Virology *351*, 466–475.

Sahin, U., Muik, A., Derhovanessian, E., Vogler, I., Kranz, L.M., Vormehr, M., Baum, A., Pascal, K., Quandt, J., and Maurus, D. (2020). Concurrent human antibody and TH1 type T-cell responses elicited by a COVID-19 RNA vaccine. medRxiv. https://doi.org/10.1101/2020.07.17.20140533.

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.

Seow, J., Graham, C., Merrick, B., Acors, S., Steel, K.J.A., Hemmings, O., O'Bryne, A., Kouphou, N., Pickering, S., Galao, R.P., 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.

Sette, A., and Crotty, S. (2020). Pre-existing immunity to SARS-CoV-2: the knowns and unknowns. Nat. Rev. Immunol. *20*, 457–458.

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.

Thieme, C., Anft, M., Paniskaki, K., Blázquez Navarro, A., Doevelaar, A., Seibert, F.S., Hoelzer, B., Konik, M.J., Brenner, T., and Tempfer, C. (2020). The SARS-CoV-2 T-cell immunity is directed against the spike, membrane, and nucleocapsid protein and associated with COVID 19 severity. J. Clin. Invest. Published online August 24, 2020. https://doi.org/10.1172/JCI140965.

Vabret, N., Britton, G.J., Gruber, C., Hegde, S., Kim, J., Kuksin, M., Levantovsky, R., Malle, L., Moreira, A., Park, M.D., et al.; Sinai Immunology Review Project (2020). Immunology of COVID-19: Current State of the Science. Immunity *52*, 910–941.

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.

Wherry, E.J., and Ahmed, R. (2004). Memory CD8 T-cell differentiation during viral infection. J. Virol. *78*, 5535–5545.

Yewdell, J.W. (2006). Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. Immunity *25*, 533–543.

Zhao, J., Zhao, J., and Perlman, S. (2010). T cell responses are required for protection from clinical disease and for virus clearance in severe acute respiratory syndrome coronavirus-infected mice. J. Virol. *84*, 9318–9325.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
APC-Cy7 anti-CD3 (HIT3a)	Biolegend	Cat#300318; RRID:AB_314054
AF647 anti-CD8 ((SK1)	Biolegend	Cat#344726; RRID:AB_2563452
BV421 anti-CD8 (SK1)	Biolegend	Cat#344748; RRID:AB_2629584
BV510 anti-CD45RA (HI100)	Biolegend	Cat#304142; RRID:AB_2561947
PE anti-CD45RO (UCHL1)	Biolegend	Cat#304244; RRID:AB_2564160
Pac-Blue anti-CD57 (HNK-1)	Biolegend	Cat#359608; RRID:AB_2562459
Anti-CD3 (OKT3)	ThermoFisher	Cat#16-0037-81; RRID:AB_468854
AF647 anti-CD69 (FN50)	Biolegend	Cat#310918; RRID:AB_528871
PE anti-CD137	Miltenyi Biotec	Order no: 130-093-476
AF488 anti-TCR$\alpha\beta$ (IP26)	Biolegend	Cat#306712; RRID:AB_528967
Chemicals, peptides, and recombinant proteins		
rh IL-2	Peprotech	Cat#200-02
rh IL-7	R&D Systems	Cat#207IL025
rh IL-15	R&D Systems	Cat#247ILB025
Peptides (see Table 1)	Genscript	Custom
PE- or APC-conjugated empty A*02:01 tetramers	Tetramer Shop	Cat#HA02-070
Critical commercial kits/assays		
Memory $CD8^+$ T cell Isolation Kit, Human	Miltenyi Biotec	Order no: 130-094-412
Annexin V MicroBead Kit	Miltenyi Biotec	Order no: 130-090-201
GeneJET Genomic DNA Purification Kit	Thermo Scientific	Cat#K0702
Single Cell V(D)J Reagent Kit (v1)	10x Genomics	Cat#1000005
Ella human IFNγ 3rd generation single-plex assay	Protein Simple	Cat#SPCKB-PS-002574
Deposited data		
SARS-CoV-2 coding sequences as of March 15, 2020	NCBI	taxid:2697049
SARS-CoV genome sequence	NCBI	NC_004718.3
HCoV 229E genome sequence	NCBI	NC_002645.1
HCoV NL63 genome sequence	NCBI	NC_005831.2
HCoV OC43 genome sequence	NCBI	NC_006213.1
HCoV HKU1 genome sequence	NCBI	NC_006577.2
Experimental models: cell lines		
HEK293T cells	ATCC	Cat#CRL-3216 RRID: CVCL_0063
Recombinant DNA		
Coronavirus peptidome library	Agilent	Custom

RESOURCE AVAILABILITY

Lead Contact

Gavin MacBeath, CSO, TScan Therapeutics, 830 Winter Street, Waltham, MA 02451, USA, gmacbeath@tscan.com

Materials Availability

Peptides generated in this study are available for research purposes upon signing a materials transfer agreement. Peptides and peptide sequences for commercial purposes (e.g., diagnostics or vaccine development) are available through license agreement.

Data and Code Availability

T-Scan screening data are available upon request. Amino acid sequences for the coronavirus peptidome library (43,420 sequences) are available upon request. T cell receptor sequences for the purposes of therapeutic development are available through license agreement.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Ethics statement

All donors provided written consent. The study was conducted in accordance with the Declaration of Helsinki (1996), approved by the Atlantic Health System Institutional Review Board (IRB) and the Ochsner Clinic Foundation IRB and registered at clinicaltrials.gov (# NCT04397900).

Sample collection design

Patients who had recovered from COVID-19 were eligible for this study. They were required to be >18 years of age and have laboratory-confirmed diagnosis of COVID-19 using CDC or state health labs or at hospitals using an FDA Emergency Use Authorized molecular assay. Time since discontinuation of isolation was required to be >14 days and discontinuation of isolation followed CDC guidelines (accessed on March 19, 2020) using either test-based or non-test-based criteria for patients either in home isolation or in isolation at hospitals. Patients were also required to have no anti-pyretic use for >17 days and be able to sign informed consent for blood draws for 4 tubes of whole blood with approximately 7.5 mL of blood per tube. Eligible patients were identified by the participating sites through advertising and direct contact. Case report forms did not contain identifying information. Samples were de-identified at the participating sites with an anonymous code assigned to each sample. Anonymized blood samples were sent to TScan laboratories with limited demographic and clinical data. Demographics included age, sex, and ethnicity. Clinical data included date of diagnosis, specifics of diagnostic testing, duration of symptoms, and whether the patient required hospitalization, supplemental oxygen, or ICU care/ ventilator support. Comorbidities and current medications were also recorded.

Patient recruitment and demographics

Convalescents who met eligibility criteria and consented to described procedures were enrolled and sampled from two sites: Atlantic Health (New Jersey, 51 samples) and Ochsner (New Orleans, 27 samples). These sites played a critical role in treating patients from early epicenters of SARS-CoV-2 outbreaks. Recruitment materials clearly requested patients that had recovered from COVID-19 with the goal of designing effective vaccines and treatments for this indication. Average self-reported duration of symptoms was 18 days (1-80 days range) in females and 21 days (0-76 days range) in males. Hospitalizations made up ~32% of total convalescent samples received, with 31% requiring oxygen and 5% placed on a ventilator.

METHOD DETAILS

Isolation of PBMCs and Memory CD8+ T cells

Blood samples were collected in four 10-mL K2 EDTA vacutainer tubes (Becton Dickinson) and processed within 24-30 h to PBMCs or memory CD8+ T cells. A 1-mL sample was removed and centrifuged at 500x*g* for 10 min to obtain plasma. To isolate PBMCs, blood samples were diluted with an equal volume of MACS separation buffer (phosphate buffered saline, 0.5% bovine serum albumin, 2 mM EDTA), then layered onto lymphocyte separation media (Corning) and centrifuged at 1200x*g* for 20 min. The interface was removed and washed once with MACS buffer before further processing or cryopreservation. Memory CD8+ T cells were isolated from PBMCs using MACS microbead kits according to the manufacturer's instructions (Miltenyi). Following separation, purity was confirmed using antibodies to CD3, CD8, CD45RA, CD45RO and CD57 (Biolegend). Immediately following isolation, memory CD8+ T cells were expanded by co-culturing with $2x10^7$ mitomycin C treated (50 μg/mL, 30 min) allogenic PBMCs in the presence of 0.1 μg/mL anti-CD3 (OKT3, eBioscience), 50 U/mL recombinant IL-2 (Peprotech), 5 ng/mL IL-7 and 5 ng/mL IL-15 (R&D Systems). After 10 days of expansion, the cells were collected and cryopreserved.

Peptide Library Generation and Cloning

Coding sequences of all deposited SARS-CoV-2 strains were downloaded from NCBI on March 15, 2020, totaling 1,117 proteins. Full-genome coding sequences from SARS-CoV (NC_004718.3), HCoV 229E (NC_002645.1), HCoV NL63 (NC_005831.2), HCoV OC43 (NC_006213.1) and HCoV HKU1 (NC_006577.2) were downloaded from NCBI. All full-length ORFs were divided into 61-aa fragments tiled every 20-aa. Protein fragments containing the 32 CMV, EBV, and Flu (CEF) peptides were added as positive controls. All protein fragments were reverse translated with ten unique nucleotide sequences each, synthesized on a releasable microarray (Agilent), and cloned into the pHAGE CMV NFlagHA DEST vector.

TScan Screen

MHC null HEK293 cells were engineered to express a fluorescent reporter activated by T cell killing. To stimulate T cells for antigen screens, $1.5x10^7$ memory CD8+ T cells were thawed and re-stimulated as described above by co-culturing with $3x10^8$ mitomycin C-treated (50 μg/mL, 30 min) allogenic PBMCs in the presence of 0.1 μg/mL anti-CD3 (OKT3, eBioscience), 50 U/mL recombinant

IL-2 (Peprotech), 5 ng/mL IL-7 and 5 ng/mL IL-15 (R&D Systems). After expansion for 7 days, the T cells were added to library-transduced reporter cells at an effector to target ratio of 1.25:1 and incubated at 37°C for 4 h. After incubation, cells were harvested by trypsinization and labeled with Annexin V magnetic microbeads (Miltenyi) according to the manufacturer's instructions. Annexin-labeled cells were isolated using an AutoMACS Pro (Miltenyi), and reporter-positive cells were sorted using a MoFlo Astrios EQ cell sorter (Beckman Coulter).

Genomic DNA (gDNA) was extracted from sorted cells using the GeneJET Genomic DNA Purification Kit (Thermo Scientific). Samples were then subjected to 2 rounds of PCR. In the first round, primers amplified the antigen cassette from the extracted gDNA. Following PCR purification using AMPure XP beads, the second round of PCR added sequencing adaptors and sample-specific index sequences to the amplicon. Samples were then purified using AMPure XP beads and pooled to equal quantities of DNA. Amplicons were sequenced on either an Illumina MiSeq or Illumina NextSeq using the standard Illumina sequencing primer. A 150-cycle kit was used for either instrument and sequencing was performed with read lengths: 110 bp- read 1, 8bp- i7 index, 8bp- i5 index.

Nucleotide sequences were mapped to individual nucleotide tiles and read counts for each library entity representing identical amino acid tiles were summed. The proportion of read counts for each tile was calculated for each screen replicate (n = 4) and for the input for each pool of transduced reporter cells, and enrichments of each tile were calculated by dividing the proportion of the tile in the screen replicate by the proportion of the tile in the input library. A modified geometric mean of the enrichment of an identical tile across the 4 screen replicates (calculated by adding 0.1 to all enrichment values and taking the geometric mean) was used to identify reproducible screen hits. Specific MHC-binding epitopes for each tile above the threshold of 2-fold enrichment were predicted using NetMHC4.0. Candidate epitopes for each tile were selected by identifying predicted strong binding epitopes shared across overlapping adjacent and redundant tiles that were enriched in the screen. To collapse data from multiple tiles into a single datapoint for each patient, the arithmetic mean of all the tiles containing the indicated epitope was calculated.

Peptide Validation Assay

$5x10^4$ monomeric MHC reporter cells were seeded into 96 well plates and rested for 16 h, then pulsed with 1 μg/mL of individual peptides (Genscript) for 1 h. Bulk isolated CD8+ memory T cells were thawed, washed with warm media, added to the plates at a 2:1 effector to target cell ratio, and incubated for 16 h. The cells were harvested by pipetting, transferred to V-bottom 96-well plates and centrifuged at 500x*g* for 2 min. The supernatant was removed and IFNγ was immediately measured using an Ella human IFNγ 3rd generation single-plex assay (Protein Simple) following the manufacturer's instructions. The remaining cell pellets were washed with FACS buffer (phosphate buffered saline, 0.5% bovine serum albumin, 2 mM EDTA) and stained with PE-conjugated anti-CD137 (Miltenyi), AF647-conjugated anti-CD69 (Biolegend), and BV421-conjugated anti-CD8 (Biolegend) antibodies and analyzed by flow cytometry (Cytoflex S, Beckman Coulter).

Tetramer Staining

MHC tetramers were generated by incubating each peptide with PE- or APC-conjugated empty A*02:01 tetramers (Tetramer Shop) at a final peptide concentration of 30 μg/mL for 30 min at room temperature. Two tetramer-peptide reagents with contrasting fluorophore conjugates were used in each stain cocktail at a dilution of 1:10 in FACS buffer. Bulk isolated memory CD8+ T cells were thawed, washed with warm media, and plated in V-bottom 96-well plates at $1x10^6$ cells/well. Cells were pelleted and resuspended in the tetramer stain cocktail and incubated at 37°C for 15 min prior to adding an AF488-conjugated anti-human TCR antibody (Biolegend) and incubating for an additional 15 min at room temperature. The stained cells were pelleted and washed three times before resuspending in a 5 μg/mL DAPI solution and analyzed by flow cytometry (Cytoflex S, Beckman Coulter). The limit of detection was defined as the mean + 2 SD of the frequency of three MHC-mismatched controls.

Single Cell TCR sequencing

Single-cell TCR-seq (scTCR-seq) libraries were prepared following the 10x Genomics Single Cell V(D)J Reagent Kit (v1) protocol. Briefly, cells were captured in droplets before undergoing reverse transcription. Following cDNA purification, cDNA was amplified (98°C for 45 s; 16 cycles of 98°C for 20 s, 67°C for 30 s, 72°C for 1 min; 72°C for 1 min). Following sample purification, 2 μL of each library was used for TCR sequence enrichment. TCR enriched libraries were subsequently fragmented, end-repaired, and amplified with indexing primers. The scTCR-seq libraries were sequenced on an Illumina NextSeq using a High Output v2.5 kit (150 cycles) with read lengths: 26 bp- read 1, 8 bp- i7 index, 98 bp- read 2. scTCR-seq reads were processed using cellranger 3.1.0. Reads were aligned to the GRCh38 reference genome and cellranger vdj was used to annotate TCR consensus sequences.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses were performed using GraphPad prism, Excel, or Python. The details of the statistical tests are displayed in the figure legends.