



OPEN

Immunodominant T-cell epitopes from the SARS-CoV-2 spike antigen reveal robust pre-existing T-cell immunity in unexposed individuals

Swapnil Mahajan^{1,3}, Vasumathi Kode^{2,3}, Keshav Bhojak^{1,3}, Coral Karunakaran^{1,3}, Kayla Lee², Malini Manoharan¹, Athulya Ramesh¹, Sudheendra HV¹, Ankita Srivastava¹, Rekha Sathian¹, Tahira Khan², Prasanna Kumar¹, Ravi Gupta¹, Papia Chakraborty²✉ & Amitabha Chaudhuri²✉

The COVID-19 pandemic has revealed a range of disease phenotypes in infected patients with asymptomatic, mild, or severe clinical outcomes, but the mechanisms that determine such variable outcomes remain unresolved. In this study, we identified immunodominant CD8 T-cell epitopes in the spike antigen using a novel TCR-binding algorithm. The predicted epitopes induced robust T-cell activation in unexposed donors demonstrating pre-existing CD4 and CD8 T-cell immunity to SARS-CoV-2 antigen. The T-cell reactivity to the predicted epitopes was higher than the Spike-S1 and S2 peptide pools in the unexposed donors. A key finding of our study is that pre-existing T-cell immunity to SARS-CoV-2 is contributed by TCRs that recognize common viral antigens such as Influenza and CMV, even though the viral epitopes lack sequence identity to the SARS-CoV-2 epitopes. This finding is in contrast to multiple published studies in which pre-existing T-cell immunity is suggested to arise from shared epitopes between SARS-CoV-2 and other common cold-causing coronaviruses. However, our findings suggest that SARS-CoV-2 reactive T-cells are likely to be present in many individuals because of prior exposure to flu and CMV viruses.

Uncovering the immunological responses to COVID-19 infection will help in designing and developing next-generation therapies and manage the treatment of critical COVID-19 patients. Many host factors associated with mild or severe disease symptoms have been reported. For example, leukopenia, exhausted CD8 T-cells, higher levels of T_H2 cytokines in serum, a high titer of neutralizing antibodies, blunted interferon response, dysregulation of the myeloid cell compartment, activated NK cells, and the size of the naive T-cell compartment is associated with critically ill patients^{1–5}. This wide range of variable factors shares a common immunological underpinning—that of a systemic dysregulation in immune homeostasis due to the failure of the host immune system to clear the virus during the early stages of the infection⁶. Animal and human studies have demonstrated that susceptibility to respiratory virus infections is associated with compromised CD8 T-cell immunity^{7–11}. A delay in the activation of CD8 T-cells and a lack of early IFN- γ production by the innate immune arm leads to an increase in viral load triggering overactivation of the innate and the adaptive arm of the immune system leading to a loss of immune homeostasis resulting in severe disease phenotype, including death. Therefore, an early wave of strong CD8 T-cell response may delay viral titer build-up, allowing rapid clearance of the virus by the immune system without perturbing immune homeostasis.

Healthy humans not exposed to COVID-19 show pre-existing CD4 and CD8 T-cell immunity to SARS-CoV-2 antigens^{12–14}. The pre-existing immunity to CD4 and CD8 T-cells was detected against structural and non-structural SARS-CoV-2 proteins by overlapping 15-mer peptide pools. The existence of a pool of SARS-CoV-2-reactive T-cells in unexposed individuals is thought to arise from coronaviruses that cause common cold^{13,15,16}. Whether pre-existing immunity provides any protection to SARS-CoV-2 infection or contributes to a faster recovery from infection remains speculative. Besides, it is unclear whether a pre-existing immunity, involving either CD4 or CD8 T-cells, or both, is required for maximal protection. Identifying robust pre-existing immunity against SARS-CoV-2 in the healthy population can be used as a measure to assess the mode of recovery and also viral spread in the global population.

¹MedGenome, Bangalore, India. ²MedGenome, Foster City, USA. ³These authors contributed equally: Swapnil Mahajan, Vasumathi Kode, Keshav Bhojak and Coral Karunakaran. ✉email: pchakra1@gmail.com; amitc@medgenome.com

In this study, we identified strong CD8 T-cell-activating epitopes from SARS-CoV-2 spike protein by a combination of epitope prediction and T-cell activation assays in healthy donors unexposed to SARS-CoV-2. The rationale for identifying epitopes that favor CD8 T-cell activation was twofold. First, robust CD8 T-cell activating epitopes can be formulated as second-generation vaccines for short and long-term protection against viral infection. Second, detection of pre-existing immunity in healthy donors using epitopes that favor CD8 T-cell activation may provide a framework to understand the complex immune responses observed in clinical settings. It may also shed light on the differences in morbidity and mortality in different population groups across the globe.

We developed a proprietary algorithm OncoPeptVAC to predict CD8 T-cell activating epitopes across the SARS-CoV-2 proteome. OncoPeptVAC predicts binding of the HLA-peptide complex to the T-cell receptor (TCR). We selected a cocktail of eleven 15-mer peptides with a broad class-I and class-II coverage and favorable TCR engagement predicted by the algorithm. The cocktail of peptides was tested for T-cell activation in healthy donors from the USA and India unexposed to COVID-19. We observed higher CD8 T-cell activation by the 11-peptide pool compared to the overlapping 15-mer peptide pools from the spike-S1 and S2 proteins. Homology analysis of the selected peptides with other coronavirus spike proteins indicated a lack of significant amino acid identity with any of the 11 peptides, suggesting engagement of one or more peptides in the pool to cross-reactive TCRs from other viruses, not particularly from a coronavirus. Bulk and single-cell TCR analysis revealed expanded clonotypes recognizing epitopes from CMV, Influenza-A, and other viruses to which most of us are exposed. Taken together, our findings support that strong pre-existing CD8 T-cell immunity in unexposed donors is contributed by cross-reactive TCRs from other viruses. Significantly, we discovered multiple immunodominant epitopes in our predicted pool of peptides that favored CD8 T-cell activation. Finally, we show that our cocktail of 11-peptides induced a robust immune response in convalescent patients demonstrating that these peptides are recognized by infected patients. Taken together, our study uncovered strong pre-existing CD8 T-cell immunity against SARS-CoV-2 using a small set of 11 epitopes that engaged cross-reactive TCRs recognizing epitopes from other viruses, not necessarily common cold viruses belonging to the coronavirus family as hypothesized by other studies. Additionally, our findings raise the possibility that many individuals carrying antigen-experienced T-cells against other viruses may be naturally protected against COVID-19 without prior SARS-CoV-2 infection.

RESULTS

Prediction of immunogenic epitopes favoring CD8 T-cell activation. A deep CNN model OncoPeptVAC was implemented to predict the immunogenicity of peptides based only on the peptide and HLA sequences. A total of 8870 immunogenic and non-immunogenic peptide-HLA pairs were obtained from the IEDB¹⁷. The BLOSUM encoding was used to represent the peptide and HLA molecules. The BLOSUM substitution scores encode evolutionary and physicochemical properties of the amino acids¹⁸. In addition, hydrophobicity indices and predicted HLA binding scores were also used to represent the peptide and HLA sequences.

OncoPeptVAC used the CNN model with multiple 2D convolutional layers combined with max-pooling to confirm the additive effect of different input features on the performance of the model. All the model versions were trained using fivefold cross-validation. The AUC of the final model was 0.87 based on a blind test dataset (Fig. 1A). The prediction algorithm showed a sensitivity of 0.64 and a specificity of 0.84 based on the score cut-off of 0.2 (Fig. 1B). By increasing the cut-off score to 0.5, the specificity could be further increased to 96.8 with a concomitant loss in sensitivity. OncoPeptVAC reduced the number of false positives significantly compared to the HLA-binding rank (Compare Fig. 1B,C) reducing the number of epitopes by 30% that needed to be screened in a T-cell activation assay to identify true immunogenic epitopes. For example, to identify 50% (119 out of 238) of the immunogenic peptide-HLA pairs present in the blind test dataset, 256 top peptide-HLA pairs from OncoPeptVAC prediction needed to be screened, compared to 753 top peptide-HLA pairs predicted by netMHCpan-4.0.

The prediction algorithm was applied to the SARS-CoV-2 proteome and screened against 23 class-I HLAs covering over 98% of the world population. A schematic of the *in silico* screening approach is shown in Fig. 1D. Briefly, 9–11-mer peptides from the SARS-CoV-2 proteome were screened for TCR-binding against 23 HLA, and peptides with OncoPeptVAC score > 0.2 were analyzed for class-I HLA binding. Peptide-HLA pairs with a high predicted binding affinity (< 1 percentile rank) were selected, their length extended to 15-mer, and screened for class-II HLA binding (See “Methods” for details). Peptides with favorable TCR binding and class-I/II HLA binding features were selected for further validation. The number of predicted immunogenic epitopes from SARS-CoV-2 protein-coding genes is shown in Fig. 1E. The distribution of OncoPeptVAC scores against different class-I HLA genes indicates a higher number of favorable TCR-binding peptides for HLA-B and C compared to HLA-A (Fig. 1F). Natural biases in HLA restrictions have been reported for immunogenic HIV epitopes¹⁹.

T-cells from unexposed donors respond to OncoPeptVAC-predicted peptides. We performed T-cell activation assay using a set of 11 prioritized epitopes from the SARS-CoV-2 spike antigen (Table 1) in unexposed donors. We selected epitopes from the spike antigen because it is highly immunogenic and generates strong B and T-cell responses. Besides, identifying strong immunogenic epitopes from the spike antigen can become useful reagents to study mechanisms of immune toxicity, and for long-term immune monitoring studies in naïve and vaccinated populations. The 15-mer peptides contained in the 11 selected epitopes cover different segments of the RBD and the non-RBD regions of the spike antigens and few peptides harbor ACE2 receptor binding sites (Fig. 2A and Table 1). Many of the predicted peptides reside in flexible regions of the spike protein that could favor efficient processing and presentation. Only two out of the 11 peptides showed 100% identity to the 315 peptides present in the spike-S1 and S2 pools obtained from a commercial vendor (Table 2) (See the “Methods” section for the creation of the peptide pool).

We screened PBMCs from 14 unexposed donors from the US collected between 2016–2018 and India (2015–2017), much before SARS-CoV-2 was recognized as a global pandemic (Table-S1). Activation of T-cells

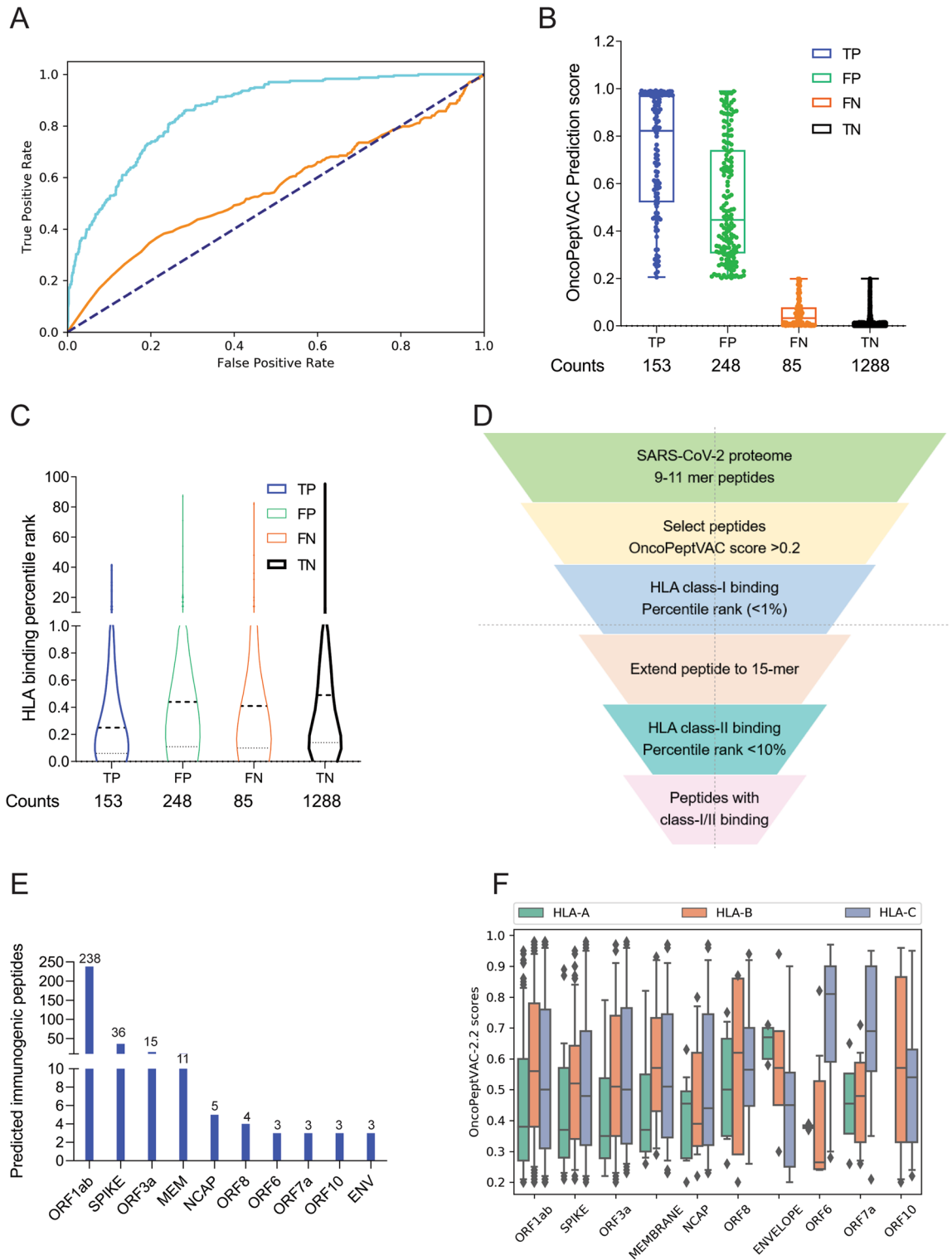


Figure 1. Identification of immunogenic epitopes from SARS-CoV-2 by OncoPeptVAC. **(A)** ROC curves of OncoPeptVAC TCR-binding and netMHCpan-4.1 HLA binding algorithms. A blind dataset of non-immunogenic or immunogenic HLA class-I binding T-cell epitopes from IEDB was used to assess the performance of OncoPeptVAC (cyan). The HLA-binding affinity of the epitopes expressed as percentile rank < 1% was used to assess the performance of netMHCpan-4.1 in predicting true immunogenic epitopes (orange). **(B)** Separation of immunogenic from non-immunogenic epitopes by OncoPeptVAC score. **(C)** Separation of immunogenic from non-immunogenic epitopes by HLA-binding percentile rank. **(D)** Schematic showing the steps used to identify immunogenic epitopes from SARS-CoV-2 proteome. **(E)** Number of immunogenic epitopes identified in different SARS-CoV-2 antigens. **(F)** HLA-A, B and C-restricted epitopes from SARS-CoV-2 proteome.

Peptide ID	# Class-I HLA restriction	#Class-II HLA restriction	OncoPeptVAC score for individual HLAs (range)	Class-I binding percentile rank (range)	Class-II binding percentile rank (range)	Presence in spike peptide pools	Other features
Pep-1	3	0	0.21–0.51	0.27–0.77	NA	Yes	One contact residue for neutralizing Ab
Pep-2	0	2	NA	NA	7.6–7.8	No	4 ACE contact residues and 2 contact residues for neutralizing Ab
Pep-3	1	3	0.27	0.83	1.5–8.5	No	
Pep-4	7	9	0.43–0.94	0.02–0.25	0.58–10.0	No	
Pep-5	2	1	0.24–0.28	0.45–0.49	3.1	Yes	4 ACE contact residues
Pep-6	6	10	0.28–0.9	0.37–0.97	0.27–9.6	No	
Pep-7	11	18	0.21–0.94	0.06–0.8	0.49–9.3	Yes	
Pep-8	14	16	0.21–0.7	0.01–0.87	0.97–10.0	No	
Pep-9	27	13	0.22–0.96	0.02–0.89	1.8–8.8	No	
Pep-10	4	9	0.23–0.83	0.19–0.71	0.45–8.8	No	5 ACE contact residues and 2 contact residues for neutralizing Ab
Pep-11	6	2	0.22–0.88	0.44–0.9	7.2–8.0	No	

Table 1. Peptides selected by OncoPeptVAC for use in T-cell activation assays.

using the cocktail of 11 peptides (All-peptide mix) was compared to the responses from Spike-S1 (157 peptides) and S2 (158 peptides) pools (see “Methods” for assay details). At 48 h, the All peptide-mix induced a strong IFN- γ response in CD8 T-cells (Fig. 2B left panel), and a weaker response in CD4 T-cells (Fig. 2C, left panel). The IFN- γ response by the All peptide-mix was stronger in CD8 T-cells, compared to responses from spike-S1 and S2 peptide pools (Fig. 2B, left panel). The 4-1BB response in both cell types was weaker at 48 h (Fig. 2B,C right panels). IFN- γ levels increased in CD8 T-cells at day-7 by the All-peptide mix compared to the Spike peptide pools (Fig. 2C, left panel). The higher expression of IFN- γ and 4-1BB in the CD4 and CD8 T-cells in certain donors at 7-day suggested de novo activation (Fig. 2D,E). At 48 h, 70% of the unexposed donors showed > 0.5% IFN- γ response in CD8 T-cells to the predicted peptide mix suggesting recall to pre-existing antigen-experienced CD8 T-cells (Fig. 2F). In most donors, the maximal response was detected by 7-days, but in donors D142 and D176 the response peaked at 48 h and declined thereafter (Fig. 2F). Although the use of 15-mer peptides is expected to skew the response towards CD4 T-cells, we observed a stronger CD8 T-cell response to the Peptide-mix suggesting that the 15-mer peptide though added exogenously, was processed and presented by class-I HLAs efficiently. Taken together, the results demonstrate that the use of OncoPeptVAC identified potent CD8 T-cell epitopes in the spike antigen that could not have been detected by using large overlapping peptide pools used in T-cell activation assays.

Next, we tested individual peptides from the mix to assess their contribution to T-cell activation. The magnitude and kinetics of IFN- γ and 4-1BB induction in CD8 T-cells by individual peptides were variable in different donors (Fig. 2G,H and Figure S1, S2). Peptide-7 was an exceptionally strong CD8 T-cell epitope inducing IFN- γ in 3 out of 7 donors (Fig. 2G,H and Supplementary Figure S1E). Most immunogenic epitopes activated CD8 T-cells at 48 h and achieved > 2% activation by 7-day. The early activation at 48 h suggested that these epitopes engaged pre-existing T-cell immunity in the unexposed donors. However, activation at day-7 could come from both pre-existing and naïve T-cells. In many donors, a strong CD4 T-cell response was detected by individual peptides (Figure S3–4). Interestingly, in many donors, both IFN- γ and 4-1BB expression was induced by multiple individual peptides, although the magnitude of response by the All-peptide mix was not always additive (compare donor responses to the All peptide mix, Fig. 2F, with responses to individual peptides, Fig. 2G,H and Figure S1 for IFN- γ and Figure S2A with S2B–H for 4-1BB) confirming that the T-cell activation potential of individual peptides could be masked, when present as a part of a larger peptide pool.

Multiple studies have reported pre-existing T-cell immunity in unexposed donors using spike peptide pools and attributed the response to T-cells recognizing epitopes from common cold-causing coronaviruses to which a large section of the global population is exposed^{12,13,15}. Homology analysis of the selected epitopes (see “Methods”) indicated that 6 out of the 11 peptides share > 67% sequence identity with SARS-CoV and only 1 (Peptide-11) out of the 11 peptides has over 70% identity with multiple coronaviruses (Table 3). Peptide-11 is in the S2 domain of the spike protein and showed \geq 1% CD8 T-cell response at 48 h in 1 out of 7 donors tested (D167, Fig. 2G). However, peptides 3, 6, 7, and 9 lacking significant identity to other coronaviruses (Table 3) showed \geq 1% CD8 T-cell activation at 48 h in at least one donor out of 7 (Fig. 2G,H and Figure S1). Peptide-7 induced high CD8 T-cell activation at 48 h in two donors (D089 and D225 (Fig. 2F and Figure S1E). Taken together, the data suggest that pre-existing T-cell immunity to these peptides may be derived from cross-reactive TCRs recognizing other viruses.

Analysis of antigen-specific CDR3s in responsive donors. To identify CDR3s amplified by individual peptides, or the All-peptide-mix, bulk TCR analysis was performed on antigen-stimulated PBMCs from donors

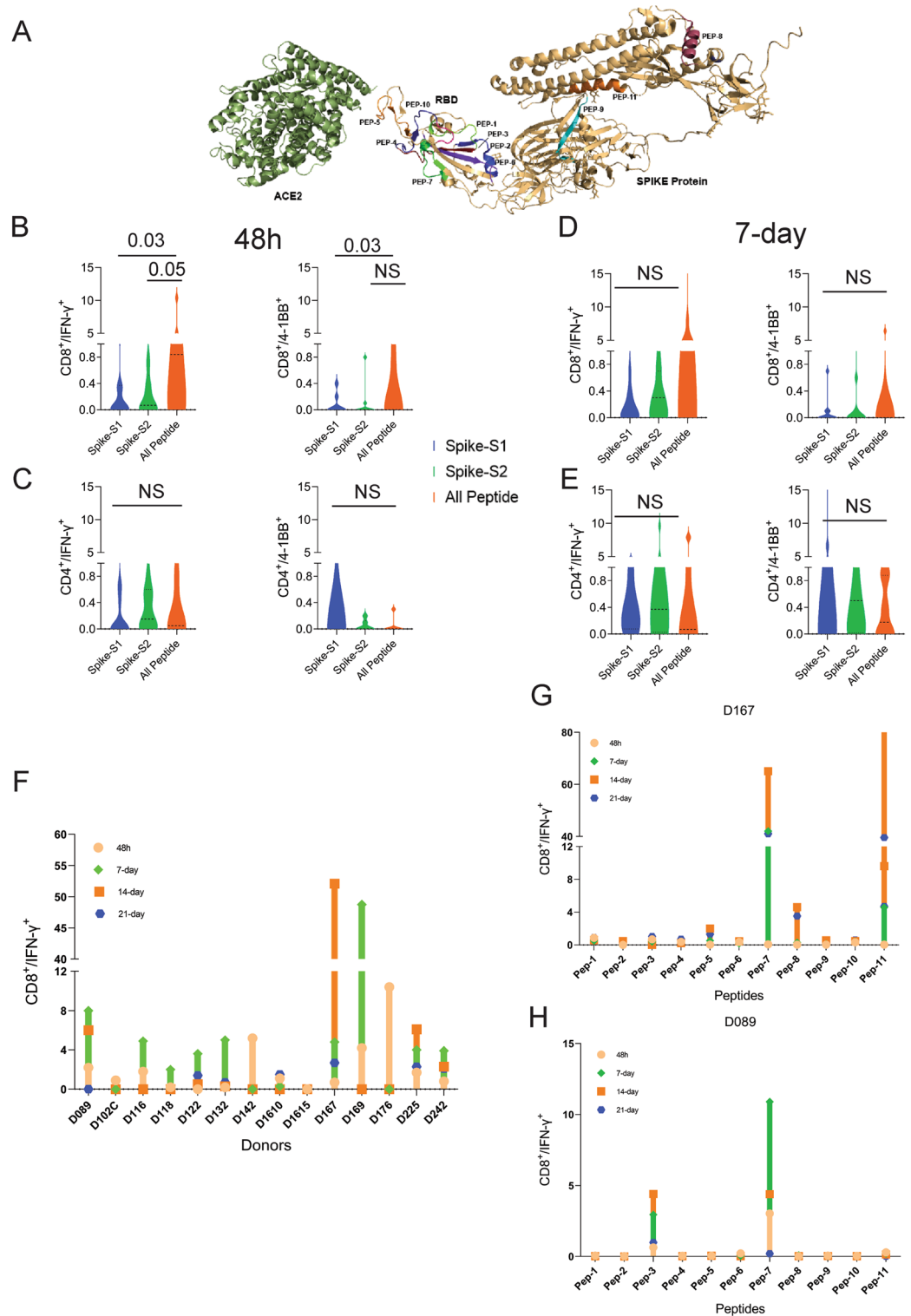


Figure 2. T-cell reactivity to SARS-CoV-2 Spike peptide pools and OncoPeptVAC prioritized peptides. Reactivity was determined by intracellular IFN- γ staining and surface expression of 4-1BB by FACS after stimulation of PBMCs from unexposed donors ($n = 14$) using separate pools of Spike-S1 and S2 peptides and the 11-Peptide-mix predicted by OncoPeptVAC (see “Methods”). (A) Structure of Spike—ACE2 receptor complex showing the location of the 11-peptides predicted by OncoPeptVAC. (B–C) T-cell activation after 48 h incubation with the peptides. (D–E) T-cell activation after a 7-day incubation with the peptides. (F) Kinetics and magnitude of CD8 T-cell activation in unexposed donor PBMCs. (G,H) CD8 T-cell response of donors D167 and D089 to individual peptides from the 11-peptide-mix. Statistical significance determined by Wilcoxon matched pairs signed rank test in Figures (B–E).

Peptide ID	OncoPeptVAC prioritized peptide	JPT peptide	Identity (%)
Pep-1	YNYKLPDDFTGCVIA	YNYKLPDDFTGCVIA	100
Pep-2	TFKCYGVSPTKLNDL	DEDDSEPVKGVKHLH	93.33
Pep-3	GCVIAWNSNNLDSKV	DEDDSEPVKGVKHLH	86.67
Pep-4	YLYRFRKSNLKPFE	DEDDSEPVKGVKHLH	86.67
Pep-5	NGVEGFNCYFPLQSY	NGVEGFNCYFPLQSY	100
Pep-6	QPYRVVLSFELLHA	DEDDSEPVKGVKHLH	93.33
Pep-7	VFNATRFASVYAWNR	VFNATRFASVYAWNR	100
Pep-8	IPFAMQMAYRFNGIG	DEDDSEPVKGVKHLH	93.33
Pep-9	YVGYLQPRFTLLKYN	DEDDSEPVKGVKHLH	93.33
Pep-10	GGNYNYLYRFRKSN	DEDDSEPVKGVKHLH	93.33
Pep-11	GKLQDVVNQNAQALN	DEDDSEPVKGVKHLH	93.33

Table 2. Identity of OncoPeptVAC prioritized peptides to the commercially available peptide pools.

Peptide ID	Peptide	MERS	229E	OC43	SARS	HKU1	NL63
Pep-1	YNYKLPDDFTGCVIA				87		
Pep-2	TFKCYGVSPTKLNDL				93		
Pep-3	GCVIAWNSNNLDSKV				62		
Pep-4	YLYRFRKSNLKPFE				53		
Pep-5	NGVEGFNCYFPLQSY						
Pep-6	QPYRVVLSFELLHA				93		
Pep-7	VFNATRFASVYAWNR				80		
Pep-8	IPFAMQMAYRFNGIG	60		47	100	47	
Pep-9	YVGYLQPRFTLLKYN	60			67		
Pep-10	GGNYNYLYRFRKSN				73		
Pep-11	GKLQDVVNQNAQALN	71	64	71	100	73	71

Table 3. Selected peptides from SARS-CoV-2 and their homology to other coronaviruses. Sequence identity of the shortlisted 15-mer peptides against homologous peptides in other coronaviruses as identified based on an E-value cut-off of 0.01 and a minimum cut-off length of 11 amino acid residues.

D089 and D225 at days 7, 14, and 21 post-stimulation with antigens (see “Methods”). Both donors showed a robust IFN- γ response to Pep-7, and the Pep-Mix, but not to Pep-1 (Figure S4). Diversity and clonal amplification of unique public and private CDR3s were analyzed at three different time points (Fig. 3A,B). Both the donors showed clonal expansion of multiple public CDR3s recognizing HCMV, human herpes virus-5 (HHV-5), and Influenza-A peptides when stimulated with Pep-7 and All-peptide mix, but not with Pep-1 (Table S2). HCMV and HHV-5 CDR3s were expanded in donor D089 (Fig. 3A, top panel), whereas D225 showed expansion of HCMV and Influenza-A CDR3s (Fig. 3A, bottom panel). Significantly, these CDR3s were not amplified by Spike-S1 and S2 peptide pools or by Pep-1, the latter failed to activate T-cells in these donors. Further, CDR3s recognizing HCMV peptide NLVPMVATV in donors 089 and 225 were different, suggesting that the same antigen engages multiple cross-reactive TCRs in different donors. Next, we analyzed private CDR3s in these two donors to identify novel SARS-CoV-2 antigen-specific CDR3 (Fig. 3B). Donor 089 showed a lack of specific amplification of private CDR3s suggesting that the robust CD8 T-cell response detected in this donor may be contributed by the amplified public CDR3s (Fig. 3B, top panel). In contrast, two private CDR3s were clonally amplified by Pep-7 and All-peptide mix in D225 suggesting that the T-cell response is derived from both public and non-public TCRs in this donor (Fig. 3B, bottom panel). Given that the TCR analysis was performed on day-7 and beyond, the amplified CDR3s may be derived from both pre-existing and naïve T-cells. A list of clonally amplified public and private CDR3s detected in the two donors is given in Table S2.

To further investigate the TCR repertoire profile of donors 089 and 225 we analyzed the VDJ gene usage in the bulk CDR3 data. In D225, two V segments TRVB2 and TRVB30, and a J gene TRBJ2-1 were significantly over-represented in Pep7 and Peptide-mix treated samples (Fig. 3C), whereas in D089 TRBV12-4 and TRBJ1-2 genes were amplified (Fig. 3D).

Single-cell gene expression and TCR profiling of activated T-cells. To characterize the phenotype and functional state of activated T-cells and reveal differences between the different treatments, we performed single-cell sequencing on a 10X platform. Single-cell transcriptomics and TCR data obtained from 3500–4500 cells identified 3000–3500 unique transcripts (see “Methods”). Using graph-based clustering of uniform manifold approximation and projection (UMAP) we captured transcriptomes of 4 distinct cell types (Fig. 4A and

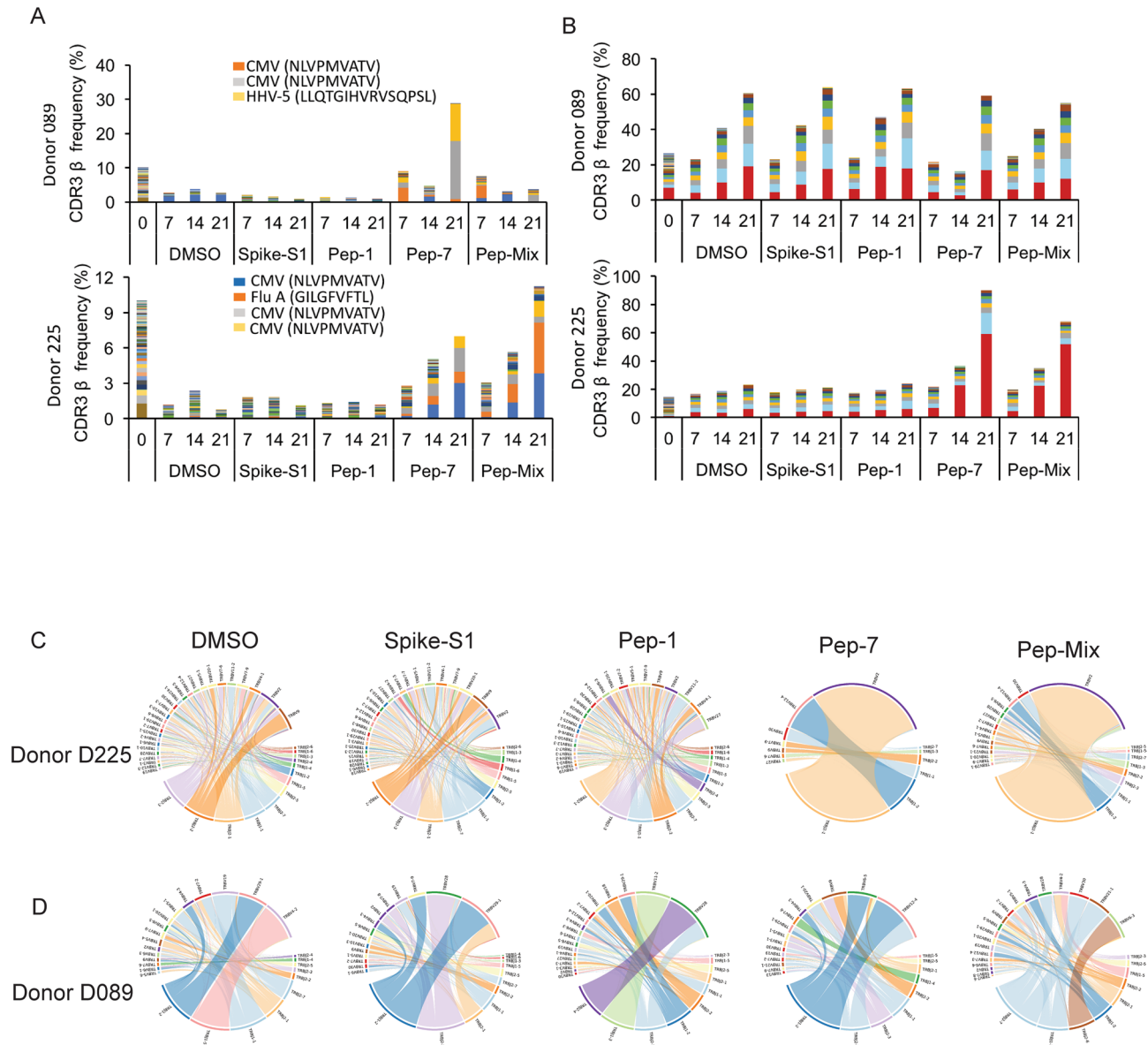


Figure 3. Bulk TCR repertoire analysis after in unexposed donors following in vitro stimulation of PBMCs at different time points with the indicated peptides. **(A)** Expanded public CDR3-βs recognizing shared antigens in the D089 (upper panel) and D225 (lower panel). **(B)** Expanded private CDR3-βs in D089 (upper panel) and D225 (lower panel). **(C)** V-J gene usage in D225. **(D)** V-J gene usage in D089.

Table S3). Our assay method is enriched for the growth and proliferation of T-cells causing depletion of other immune cell types present in PBMC in a 14-day culture. Three cell types, CD8, γ/δ , and NK-T were detected in all the samples. Compared to DMSO and Pep-7 in which the CD8 T-cell fraction was $\sim 60\%$, in spike-S1 and spike-S2 the CD8 T-cell fraction was 50% and 38% respectively. Conversely, the CD4 cluster was expanded in spike-S1 (12%) and S2 (27%) compared to DMSO (7%) suggesting that the spike peptide pools engaged CD4 T-cells (Table S3). The single-cell transcriptomic analysis further revealed that Pep-7 induced effector phenotype in the CD8 T-cell cluster by the expression of activation markers IFN- γ , 4-1BB (Fig. 4B) TNFRSF9, FAS, and TIGIT (Compare Figure S5D with S5A-C). The top 10 Pep-7-expanded clonotypes were CD27⁺/SELL⁻ suggesting transition towards effector memory phenotype (Fig. 4B). Spike-S1 and S2 peptides induced CD27⁺/SELL⁻ T-cells in the CD4 clones 6 and 4 respectively (Fig. 4B). Single-cell data revealed amplification of TRBV2 (40%) and TRBJ2-1 (32%) in Pep-7 stimulated T-cells (Fig. 4D) confirming the results from the bulk TCR analysis (Fig. 3C,D).

Next, we mapped CDR3-β to specific clones from each treatment (Fig. 4C). The DMSO, spike-S1, and S2-treated samples shared many clonotypes among themselves in the same frequency range suggesting weak antigen-induced activation and proliferation of T-cells. However, Pep-7 treated sample was enriched in many CDR3-β clones absent in other samples indicating the specificity of response (Fig. 4C, red bars). Four clones among the top-20 clones encoded CMV and flu-specific CDR3-βs in Pep-7 treated sample but not in other

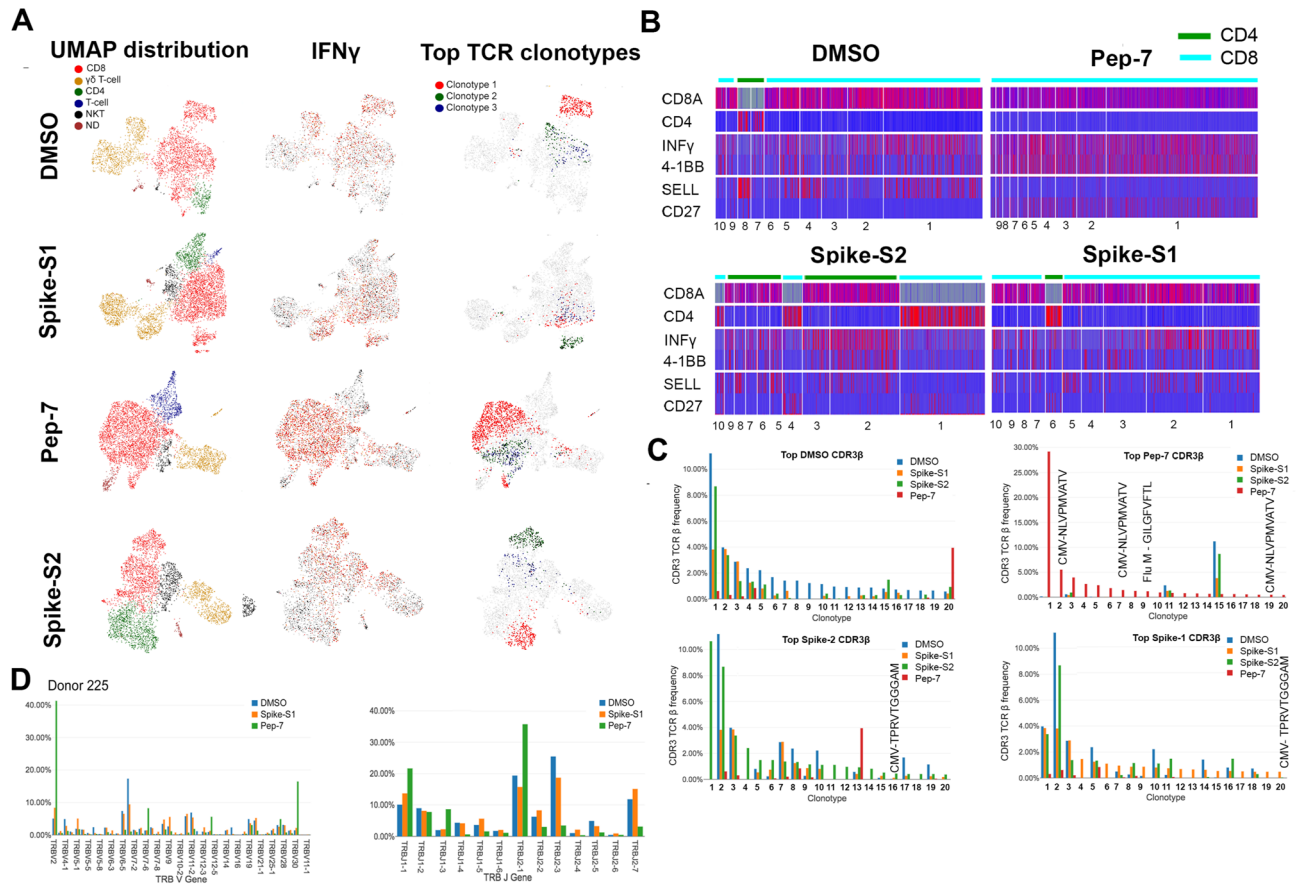


Figure 4. UMAP projection of different cell types identified in unexposed donor D225 after 14 days of in vitro stimulation assay with different antigens. (A) Clusters of different cell types and their relative proportions present in the assay mixture (left panel). Clusters expressing IFN- γ (middle panel) and the top-3 amplified clonotypes (right panel). (B) Heat map showing the expression of cell-type and cell-phenotype-specific markers in the top-10 amplified TCR- β clones. (C) Frequency of CDR3- β recognizing public and private antigens in the top 20 clonotypes. (D) Amplified V and J-genes in the top-20 clonotypes.

samples confirming the findings from the bulk TCR data that the peptide engaged cross-reactive TCRs (Fig. 4C). The expanded CDR3- β detected in the Spike-S2 treated sample belonged to CD4 T-cells (Fig. 4B, Spike-S2 panel, clonotype-1). Taken together, single-cell TCR analysis demonstrated that the immunogenic SARS-CoV-2 epitope engaged many unique CDR3- β s not shared by spike-1 and spike-2 peptide pools.

The bulk and the single-cell TCR analyses demonstrated that the SARS-CoV-2 epitope identified in this study engaged both cross-reactive public CDR3s and unique CDR3s not associated with known viral antigens and favored specific V-J gene usage. Further, the expansion of TRBV-2 and TRBJ2.1 by Pep-7 and by the Pep-mix in D225 confirmed that out of the 11-peptides contained in the Pep-mix, Pep-7 contributed to all of the T-cell responses observed in this donor.

Antigen-specific clonal expansion and T-cell phenotype. Next, we analyzed the clonal composition and phenotype of T-cells to investigate the dynamics of antigen-specific T-cell response in the treated samples. We analyzed the top-30 clones for their phenotype by the expression of 25 marker genes (Figure S6). In all samples, including DMSO, CD8 T-cell clonotypes were more frequent (Figure S6A-D). As expected, the CD4 T-cell compartment was expanded in Spike-S1 and S2-treated samples (20 and 25% of all clonotypes respectively) compared to DMSO (6.5%) (Figure S6B-C). The CD4 T-cells expressed TNFSF4 (OX-40) suggesting activation, although they failed to express IFN- γ (Figure S6B-C). A few expanded CD4 clones in the Spike-S1 and S2 treated samples showed a high expression of IL17RB suggesting polarization towards a T_H17 phenotype (Figure S6B-C). In the Pep-7 treated sample, almost all clonotypes in the top-30 were CD8 T-cells. The highly expanded clones expressed multiple T-cell activation markers (Figure S6D). Interestingly, in addition to the activation markers, these cells expressed higher levels of IL2RA (CD25) suggesting differentiation towards an effector memory phenotype (Figure S6D). CD25 expression was low in the CD8 T-cell compartment in other samples. Taken together, the results of the transcriptomic analysis highlighted that the strong immunogenic CD8 T-cell epitope identified in this study preferentially engaged CD8 T-cells pushing them towards an effector and effector memory phenotype. The spike-S1 and S2 peptide pools on the other hand engaged both CD8 and CD4 T-cells and modulated the CD4 T-cells towards a T_H17 phenotype.

Response of convalescent COVID-19 patients to predicted epitopes. To assess whether the predicted epitopes are recognized by COVID-19 infected patient T-cells, we tested the spike-S1, S2, and All-peptide mix on seven asymptomatic, five with mild-moderate symptoms, and five severe convalescent patients requiring ICU admission (Table S4) and analyzed their CD4 and CD8 T-cell response after 48 h. There was a significant donor to donor variation in T-cell response to all the antigens, however, patients experiencing mild to moderate symptoms exhibited slightly higher induction of IFN- γ in CD8 T-cells (Fig. 5A,B and Figure S13). Asymptomatic (AS) and severe patients showed weaker CD8 and CD4 T-cell response to all antigens (Fig. 5) Spike-S2 peptide pool induced stronger 4-1BB induction in CD8 and CD4 T-cells compared to the All-peptide mix (Fig. 5B,D). Taken together, our results confirm that the epitopes prioritized by the algorithm were recognized by COVID-19 infected patient T-cells, and the IFN- γ response induced by the All-peptide mix was skewed towards CD8 T-cells in line with our assay results on unexposed donors (Figure S13).

Discussion

A wide array of respiratory viruses induces severe pneumonia, bronchitis, and even death following infection. Despite the immense clinical burden, there is a lack of efficacious vaccines with long-term therapeutic benefits. Most current vaccination strategies employ the generation of broadly neutralizing antibodies, however, the mucosal antibody response to many respiratory viruses is short-lived and declines with age. In contrast, several studies on respiratory viruses have shown the presence of robust virus-specific CD8-T cell responses which have been shown to last for decades. Therefore, vaccine designs for emerging respiratory viruses need consideration and rational inclusion of CD8 epitopes to confer long-term resistance²⁰.

This study demonstrates the existence of strong CD8 T-cell activating epitopes in the spike antigen and uncovers robust pre-existing CD8 T-cell immunity in unexposed donors. Several studies have reported pre-existing T-cell immunity in unexposed donors and attributed these to infections by common cold-causing human coronaviruses^{12,13,15}. Other studies, on the contrary, have reported a lack of pre-existing T-cell immunity in unexposed donors^{21,22}. These differences can arise from the composition of peptide pools used in the assay since each group employed different selection strategies, the number of epitopes used by different groups was variable, differences among donor HLAs, dominant V and J genes in donors, and the assay method. By using a smaller number of epitopes, and donors from two different regions of the globe, the USA and India, our findings confirm the existence of robust T-cell immunity in unexposed donors. Our results differ from other studies in two important aspects. First, published studies thus far have reported robust CD4 T-cell responses and a relatively weaker CD8 T-cell response in both unexposed and convalescent subjects, whereas we show strong CD8 T-cell response in both unexposed donors and convalescent patients. In fact, in our assays the IFN- γ response was significantly higher at 48 h than reported in other studies. Although, the assay conditions—such as the use of 15-mer peptides and the combination of cytokines used in the published studies could have favored a CD4 response over a CD8 response¹³, however, we suspect that the epitope selection strategies and the use of large peptide pools by other groups may have masked the detection of strong CD8 T-cell epitopes. A second novel aspect of our study is the demonstration that the selected CD8 T-cell epitopes engaged cross-reactive TCRs in unexposed donors to mount a strong T-cell response. This finding, which needs to be validated using a larger pool of donors across different ethnic populations has significant implications in COVID-19 vaccine development efforts²³ and the spread of the infection in different regions of the world²⁴.

To identify strong CD8 T-cell epitopes, we developed a novel TCR-binding algorithm OncoPeptVAC that selects epitopes favorable for TCR-binding. In all epitope screening methods, epitope selection is primarily based on class-I and II HLA-binding affinity, which predicts surface presentation of antigen in complex with HLA²⁵, but not the interaction of the peptide-HLA complex with a TCR²⁶. By incorporating features that predict TCR-binding of a peptide, our algorithm OncoPeptVAC successfully identified many CD8 T-cell epitopes in a small pool of 11 peptides used in T-cell activation assays. The TCR-binding algorithm is especially suitable for reducing the number of epitopes that need to be screened to identify robust CD8 T-cell activating epitopes. For example, our algorithm predicted 83 peptides from all SARS-CoV-2 proteins excluding ORF1, which is a much smaller number compared to the number of peptides screened in some of the published studies to identify pre-existing T-cell immunity^{12,13,21,27}. A second factor that may have resulted in the identification of strong CD8 T-cell activating epitopes is the avoidance of epitope competition. Using a large pool of peptides to screen for T-cell responses ensures broad coverage of all HLAs, but has the disadvantage that strong immunogenic epitopes are not detected efficiently. Some of the peptides predicted by our algorithm produced > 5% CD8 T-cell response in healthy donors by 14-days. In the same donors, the response from spike-S1 and S2 peptide pools containing 157 and 158 peptides respectively was much weaker. A similar finding was reported by Mateus et al. where deconvolution of peptide pools identified a single peptide that evoked a fivefold higher T-cell response compared to the pool¹³. Also, important to note, that the strategy of using 15-mer peptides with overlapping 10 or 11-mer sequences may not identify immunodominant epitopes. For example, out of the 11-peptides tested in our assay, only three peptides were present in the spike peptide pools.

By using a smaller pool of immunodominant CD8 T-cell epitopes, our study uncovered a fundamental feature of the host immune response to SARS-CoV-2—the existence of cross-reactive TCRs to viruses, such as CMV and Influenza that recognizes SARS-CoV-2 antigen. An early and robust T-cell response is driven by the size and the diversity of the TCR repertoire to a given antigen²⁸. The Pep-7 epitope derived from the RBD domain of SARS-CoV-2 spike antigen lacking homology to other coronaviruses expanded multiple public CDR3s recognizing immunodominant CMV epitope NLVPMVATV and Influenza epitope GILGFVFTL. Further, TCR analysis demonstrated that although a donor's TCR repertoire contains many CMV-epitope-specific CDR3s, only a few are expanded in the presence of the SARS-CoV-2 peptide. For example, donor D225 TCR repertoire has 159 CMV and 249 Influenza CDR3s of which three and one were expanded respectively. Similarly, donor

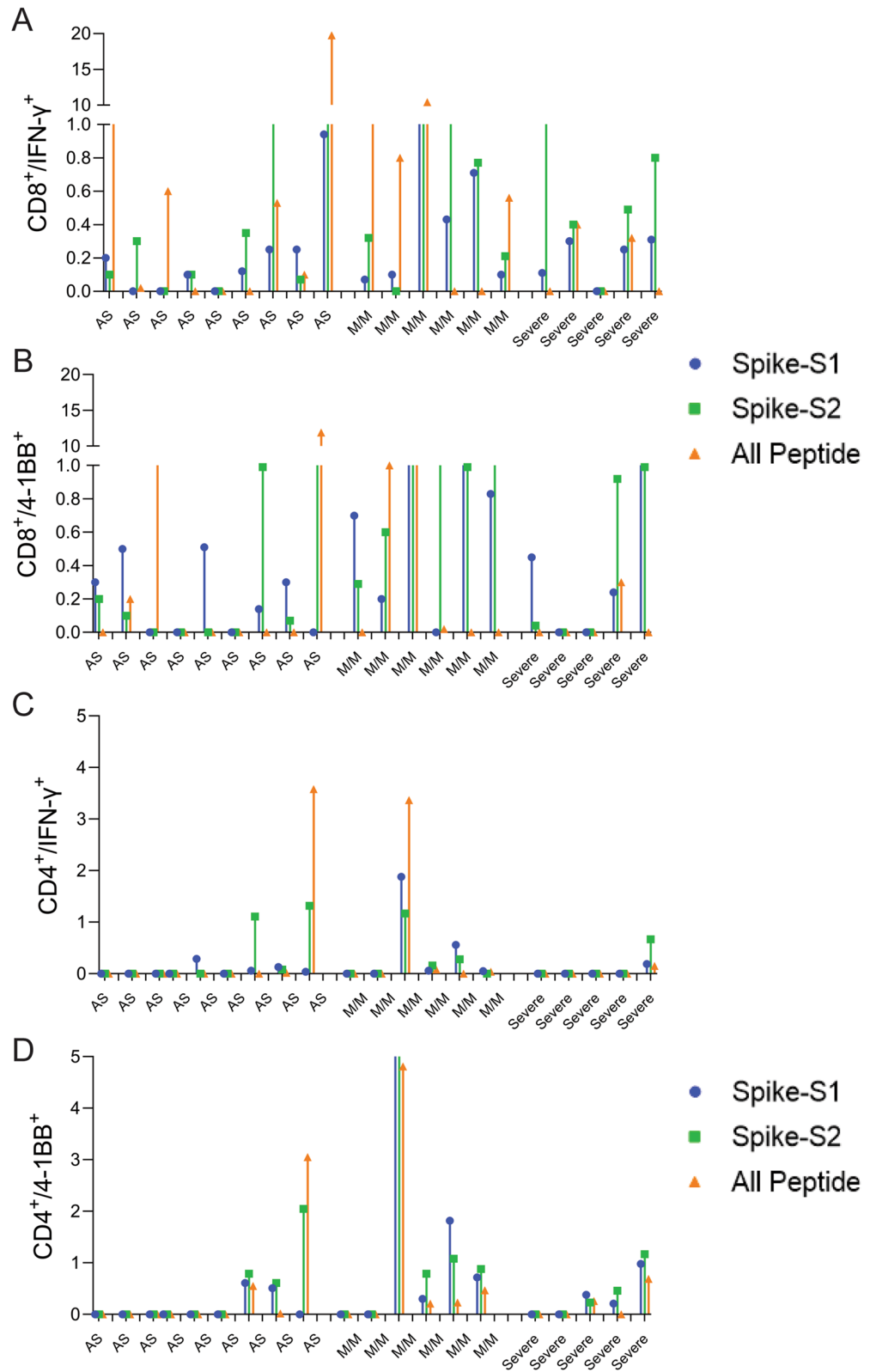


Figure 5. T-cell reactivity to Spike-S1, S2 pools, and 11-peptide-mix in asymptomatic, mild-moderate, and severe disease patients after in vitro stimulation for 48 h. **(A)** Convalescent patient PBMC collected between 45–60 days after PCR confirmation of COVID-19 infection was incubated with spike-S1 or S2 peptide pools or the All-peptide mix. T-cell activation was quantitated by intracellular IFN- γ and 4-1BB expression by FACS. T-cell activation for individual donors is plotted. **(A,B)** IFN- γ and 4-1BB expression in activated CD8 T-cells in asymptomatic (AS, 9 patients), mild/moderate (M/M, 6 patients) and severe (5 patients). **(C,D)** IFN- γ and 4-1BB expression in activated CD4 T-cells in the three patient groups. Differences in T-cell activation between groups were not statistically significant by Students T-test. Group-wise aggregated data is shown in Figure S13.

089 carries 103 NLVPMVATV specific CDR3 of which only two expanded. These findings suggest a high degree of specificity of interaction between the cross-reactive CDR3s and specific peptides from SARS-CoV-2. Significantly, the expanded CDR3s in the two donors D089 and D225 were different, even though they recognized the same CMV peptides. It has been documented that conserved features within CDR3- β allow recognition of the same pHLA complex within a group of diverse CDR3s²⁹. A robust antigen-specific T-cell response utilizes a broad range of TCRs and for many viral infections, TCR usage diversity has been positively linked to disease outcomes^{30–32}. A diverse repertoire not only allows increased structural capacity to recognize variant epitopes³³, but increases the chances that high-affinity TCRs may be present in an individual³⁴. A recent large-scale study mapped a few immunogenic regions in the SARS-CoV-2 proteome responsible for expanding many unique TCRs in a large number of convalescent COVID-19 patients and unexposed healthy donors²⁷. Immunodominant epitopes reported in our study cover some of the “hotspot” regions identified by this large-scale study²⁷. Efforts to identify cross-reactive TCRs recognizing different antigens from diverse infectious organisms can lead to the development of broad-spectrum TCR-based therapeutics against infectious diseases.

We compared the All-peptide mix with spike-1 and 2 peptide pools on a small number of convalescent patients and identified a slightly higher CD8/IFN- γ response by the peptide mix in mild to moderate disease, compared to patients with asymptomatic or severe disease. Many studies have indicated that short and long-term protection against respiratory viruses requires CD8 T-cell immunity and antibody response alone is not sufficient^{35,36}. In line with this observation, low plasma titers of neutralizing antibodies are detected in a large fraction of convalescent patients suggesting additional immune protective mechanisms, besides viral neutralization³⁷. On the contrary, high levels of neutralizing antibodies were associated with severe disease and ICU visits in many COVID-19 patients suggesting an imbalanced CD4 T-cell response is not optimal for protection^{38–40}. It has been challenging to demonstrate a strong CD8 T-cell response in COVID-19 patients in many studies. However, our findings along with a recent report from Peng et al. showed that a higher CD8 T-cell response correlated with a mild disease compared to patients with severe disease²¹.

We acknowledge several limitations of this study. First, the T-cell activation assays were performed by adding 15-mer peptides from outside, bypassing the natural processing and presentation of the antigen. Although, activation of CD8 T-cells by 15-mer peptides rely on internalization and processing for class-I presentation, the processing and the antigen presentation pathways remain poorly characterized. A more direct assessment of the existence of cross-reactive T-cells can be made by the use of spike protein-expressing antigen-presenting cells, where proteasomal processing and presentation are preserved. Second, we have not confirmed that the CMV and flu-specific TCRs amplified by exposure to SARS-CoV-2 peptides indeed recognize Pep-7 peptide. Cloning TCRs and demonstrating antigen-specificity will provide further support that protective immunity can arise from cross-reactive TCRs recognizing multiple epitopes lacking sequence identity.

In conclusion, our study demonstrates strong pre-existing CD8 T-cell immunity in many unexposed donors contributed by the engagement of cross-reactive TCRs against common CMV and flu antigens. The presence of high-quality cross-reactive TCRs can protect individuals by mounting an early CD8 T-cell response and clearing the virus. Identifying additional immunodominant epitopes in SARS-CoV-2 and their cognate TCRs can become a powerful immune monitoring tool for assessing protective immunity against SARS-CoV-2 in the population.

Methods

T-cell epitope prediction. *Dataset.* Data on 371,865 T cell assays were collected from the IEDB¹⁷. There were 105,673 CD8 T cell assays in total with 61,968 CD8 T cell assays with humans as a host. The CD8 T cell assays with HLA allele names and peptide lengths ranging 8–14 residues were further selected. HLA supertypes were replaced with their representative allele names, for example, HLA-A2 was replaced with HLA-A*02:01. The immunogenic peptide-HLA pairs tested on at least three donors with 100% response frequencies or at least tested on 5 donors with greater than 50% response frequencies were labeled as positive. The non-immunogenic peptide-HLA pairs tested on at least 3 donors with 0% response frequency were labeled as negative.

The final dataset contained 8,870 unique peptide-HLA pairs which were split randomly into 80% training and 20% test datasets. The training dataset had 884 immunogenic and 6212 non-immunogenic peptide HLA pairs. The test dataset had 238 and 1536 immunogenic and non-immunogenic peptide-HLA pairs, respectively.

Model. A Deep Convolutional Neural Network (CNN) was implemented to predict the immunogenicity of the peptide-HLA pair (provisional patent pending). The HLA alleles were represented as pseudo-sequences described as 34 amino acid residues⁴¹. The peptide and HLA pseudo-sequences were converted to the two-dimensional (2D) feature matrices of 14 \times 20 and 34 \times 20 dimensions using BLOSUM encoding¹⁸ respectively. Peptide sequences shorter than 14 residues were padded by zeroes to maintain 14 \times 20 feature matrix dimensions. Peptide sequences were also encoded into 1 \times 14 feature vector using the Kyte-Doolittle hydrophobicity scale⁴². The HLA binding percentile ranks and scores for each peptide-HLA pair were obtained using NetMHCpan-4.1⁴³ and were appended to the Kyte-Doolittle hydrophobicity scale feature vector.

The peptide and HLA feature vectors were each processed by multiple 2D convolutional filters of two different sizes followed by max-pooling layers of the same sizes serially. The peptide and HLA max-pooled layers were concatenated and processed again with multiple 2D convolutional filters followed by max-pooling layers. The 2 max-pooled layers were flattened- concatenated and then connected to a dense layer. The output of the peptide and HLA dense layer was concatenated with the hydrophobicity and HLA binding feature vector and again connected to two dense layers. The final output of the dense layer was connected to the output neuron.

Three different versions of the CNN models were trained to evaluate if the hydrophobicity scale and HLA binding scores added to the performance of BLOSUM encoding. The first version, called OncoPeptVAC-2.0, was trained only using the BLOSUM encoding. The second version, called OncoPeptVAC-2.1, was trained using

BLOSUM encoding and hydrophobicity indices. The final model version, called OncoPeptVAC-2.2, was trained using all 3 features, namely BLOSUM encoding, hydrophobicity indices, and HLA binding scores. The hyperparameters of each model version were tuned based on model performance on the blind test dataset.

The CNN was trained using fivefold cross-validation with the training dataset exclusively. The test dataset was solely used for model performance evaluation. Model performance was evaluated using AUC (area under ROC Curve) where an AUC of 0.5 represents random predictions and an AUC of 1.0 represents the perfect predictions. The TensorFlow library from Python programming language was used to implement the models.

Homology analysis. Full-length shortlisted peptide sequences of SARS-CoV-2 were blasted against the spike proteins of other coronaviruses, OC43, NL63, 229E, and HKU1. An E-value cutoff of 0.01 was used with a minimum cutoff of 11 amino acid residues was used to identify homologous peptides.

Peptides used for the T-cell activation assay. The spike-S1 and S2 peptide pools were purchased from JPT (cat# PM-WCPV-S1 and S2). The S1 and S2 pools contain 157 and 158, 15-mer peptides respectively with 10 amino acid overlap covering the full-length spike protein. For predicting immunogenic peptides from the spike protein using OncoPeptVAC, we created an in silico library of 9-mer peptides covering the S1 and S2 domains of the spike protein with an overlap of one amino acid. The 9-mer peptides prioritized by the algorithm were extended to 15-mer by adding three amino acids to either end of the peptide and assessed for class-II binding. Peptides carrying class-I and II epitopes were ranked based on their immunogenicity scores and binding to a maximum number of class-I and II HLAs.

T-cell activation assay. Unexposed donor PBMCs were obtained from the US and India for this study. PBMCs from the US were collected between 2016–2018 and purchased from Stemcell Technologies, Canada. PBMCs from India were collected between 2015–2018. COVID-19 convalescent patient blood from the US was purchased from PPA Research (USA) and the Indian samples were collected through hospitals. All participants in this study provided informed consent in accordance with protocols approved by the MedGenome Ethics Committee, MedGenome Labs, Bangalore. PBMCs were thawed, counted, and analyzed using the diagnostic panel of antibodies (Table S5). PBMCs were rested overnight in RPMI containing 10% human serum (Table S5). For T-cell activation assays, 750,000 PBMCs were incubated either with DMSO (negative control) or with different peptide pools in 0.5 ml RPMI (Gibco) + 10% Human AB serum (Sigma) + 10 ng/ml IL-15 and 10 IU of IL-2 (Stemcell Technologies, Canada). The culture media was replenished every three days with fresh media containing 10 IU of IL-2 and 10 ng/ml IL-15. On days 7, 14, and 21 of incubation, fresh peptides were added to the culture. For intracellular cytokine staining, cells were treated with Brefeldin A (BD Biosciences) for 5 h, fixed and permeabilized using BD Lysis solution and Perm2 solutions respectively followed by staining with T-cell activation panel of antibodies (Table S5). Stained cells were analyzed in BD Accuri C6 Plus to detect the expression of activation markers IFN- γ and CD137 (4-1BB) on CD4 and CD8 T cells. Data were analyzed using BD Accuri C6 Plus software.

The following Gating strategy was used: Live cells > CD3⁺ > CD4⁺ OR CD8⁺ > CD4⁺/IFN- γ ⁺ /CD4⁺/4-1BB⁺ OR CD8⁺/IFN- γ ⁺ /CD8⁺/4-1BB⁺. Representative FACS plots showing activation of IFN- γ and 4-1BB in CD4 and CD8 T-cells are given in Supplementary Figures S9–S12.

TCR sequencing and data analysis. 200,000 PBMCs was removed after 48 h, 7, 14, and 21 days from the T-cell activation assay and processed for bulk TCR sequencing.

Bulk TCR sequencing. TCR repertoire profiling was performed using the SMARTer TCR α/β Profiling Kit (Takara Bio, USA) according to the manufacturer's protocol. RNA was isolated using the Qiagen RNA isolation kit. 10 ng RNA from antigen-induced PBMCs was used as the starting material. The kit uses SMART technology (Switching Mechanism At 5'end of RNA Template) with 5'RACE to capture the entire V(D)J variable regions of TCR transcripts followed by two rounds of semi-nested PCR to obtain TCR- α and the β -chain. Libraries are prepared analyzed for quality and quantity. Sequencing was performed on a MiSeq using the 2*300 bp Reagent Kit v3 (Illumina, Inc.).

Single-cell TCR sequencing and transcriptome profiling. Approximately 10,000 cells from this assay were collected, washed, and subjected to single-cell sequencing with immune profiling to determine the gene expression profile in combination with the TCR repertoire as per the manufacturer's instruction (10 \times Genomics, CA). Sequencing was performed on Illumina NovaSeq 6000 instrument at a depth of > 50,000 reads per cell for 5' gene expression and a depth of > 5000 reads per cell for V(D)J enriched libraries. Sequencing results were evaluated using Loupe Cell and Loupe V(D)J Browsers (10 \times Genomics, CA) to assess antigen-specific CD8 T cell clonotype induction and their corresponding functional gene expression profiles.

For each sample, raw gene expression matrices were generated by Cell Ranger (v.3.0.2) coupled with the human reference version GRCh38. The gene expression data were analyzed by R software (v.3.4.4)⁴⁴ with the Seurat package (2.3.4). In brief, Low-quality cells were removed if they met one of the following criteria: > 75,000 unique molecular identifiers (UMIs); < 500 or > 7,500 genes; > 10% UMIs derived from the mitochondrial genome; > 50% of transcripts contributed by top 50 genes (Figure S7). After removing low-quality cells, normalized gene expression matrices were generated using the Seurat package. by the ,Next, the expression of the S-phase and the G2-M phase genes were used to calculate the cell cycle score for all the cells using the CellCycle-Scoring feature in Seurat. Unbiased clustering was achieved by regressing out the expression of cell cycle genes,

mitochondrial %, and number UMI from the features (Figure S8). The dimensionality of the datasets was reduced and scaled before cell clustering. Cells were clustered and annotated. Details of the Seurat analyses workflow are given in (https://satijalab.org/seurat/v2.4/pbmc3k_tutorial.html). Refer to the Supplementary Methods for codes used in single cell data analysis and for generating the figures.

Cluster annotation and differential expression of genes. After nonlinear dimensional reduction and projection of all cells into two-dimensional space by UMAP, cells were clustered together according to common features. Clusters were then classified and annotated based on expressions of canonical markers of particular cell types. Differential gene expression was performed in Seurat with default parameters. We selected top-25 upregulated DEGs with a maximum FDR value of 0.01 and annotated the clusters based on the expression of these upregulated genes. The heatmap and dot plots were generated using the DoHeatmap and DotPlot function in Seurat. Refer to the Supplementary Methods for codes used in single cell data analysis and for generating the figures.

TCR V(D)J sequencing and analysis. Full-length TCR V(D)J segments were enriched using a Chromium Single-Cell V(D)J Enrichment kit according to the manufacturer's protocol (10X Genomics). Demultiplexing, gene quantification, and TCR clonotype assignment were performed using Cell Ranger (v.3.0.2) vj pipeline with GRCh38 as reference. TCR diversity metric, containing clonotype frequency and barcode information, was obtained. Cells with at least one productive TCR α -chain (TRA) and one productive TCR β -chain (TRB) were retained for further analysis. Each unique TRA(s)-TRB(s) pair of TRA-TRB was defined as a clonotype. The presence of identical clonotypes at least in two cells was considered to be clonal, and the number of cells containing the same TRA-TRB pairs defined clonal amplification of a clonotype. Using barcode information, TCR clonotypes were projected on UMAP and Dot plots. Public TCRs were mapped to the IEDB and VDJdb annotated databases using the TRB sequence. Refer to the Supplementary Methods for codes used in single cell data analysis and for generating the figures.

Data availability

Bulk TCR sequencing and single cell sequencing data will be made available upon request to Amitabha Chaudhuri.

Received: 19 November 2020; Accepted: 4 June 2021

Published online: 23 June 2021

References

- Tay, M. Z., Poh, C. M., Renia, L., MacAry, P. A. & Ng, L. F. P. The trinity of COVID-19: Immunity, inflammation and intervention. *Nat. Rev. Immunol.* **20**(6), 363–374 (2020).
- Wang, D. *et al.* Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* **2**, 2 (2020).
- Merad, M. & Martin, J. C. Pathological inflammation in patients with COVID-19: A key role for monocytes and macrophages. *Nat. Rev. Immunol.* **20**(6), 355–362 (2020).
- Arunachalam, P. S. *et al.* Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* **369**(6508), 1210–1220 (2020).
- Ballou, M. & Haga, C. L. Why do some people develop serious COVID-19 disease after infection, while others only exhibit mild symptoms? *J. Allergy Clin. Immunol. Pract.* **2**, 2 (2021).
- Blanco-Melo, D. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*. 2020;Pre-print.
- Schmidt, M. E. & Varga, S. M. The CD8 T cell response to respiratory virus infections. *Front. Immunol.* **9**, 678 (2018).
- Taylor, P. M. & Askonas, B. A. Influenza nucleoprotein-specific cytotoxic T-cell clones are protective in vivo. *Immunology* **58**(3), 417–420 (1986).
- Askonas, B. A., Taylor, P. M. & Esquivel, F. Cytotoxic T cells in influenza infection. *Ann. N. Y. Acad. Sci.* **532**, 230–237 (1988).
- Graham, B. S., Bunton, L. A., Wright, P. F. & Karzon, D. T. Role of T lymphocyte subsets in the pathogenesis of primary infection and rechallenge with respiratory syncytial virus in mice. *J. Clin. Invest.* **88**(3), 1026–1033 (1991).
- Hall, C. B. *et al.* Respiratory syncytial viral infection in children with compromised immune function. *N. Engl. J. Med.* **315**(2), 77–81 (1986).
- Grifoni, A. *et al.* Targets of T Cell Responses to SARS-CoV-2 Coronavirus in Humans with COVID-19 Disease and Unexposed Individuals. *Cell* **181**(7), 1489–1501 (2020).
- Mateus, J. *et al.* Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* **2**, 2 (2020).
- Weiskopf, D. *et al.* Phenotype and kinetics of SARS-CoV-2-specific T cells in COVID-19 patients with acute respiratory distress syndrome. *Sci. Immunol.* **5**(48), 2 (2020).
- Braun, J. *et al.* SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* **2**, 2 (2020).
- Lipsitch, M., Grad, Y. H., Sette, A. & Crotty, S. Cross-reactive memory T cells and herd immunity to SARS-CoV-2. *Nat. Rev. Immunol.* **20**(11), 709–713 (2020).
- Vita, R. *et al.* The immune epitope database (IEDB): 2018 update. *Nucleic Acids Res.* **47**(D1), D339–D343 (2019).
- Nielsen, M. *et al.* Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein Sci.* **12**(5), 1007–1017 (2003).
- Arora, J. *et al.* HIV peptidome-wide association study reveals patient-specific epitope repertoires associated with HIV control. *Proc. Natl. Acad. Sci. U.S.A.* **116**(3), 944–949 (2019).
- Channappanavar, R., Zhao, J. & Perlman, S. T cell-mediated immune response to respiratory coronaviruses. *Immunol. Res.* **59**(1–3), 118–128 (2014).
- Peng, Y. *et al.* Broad and strong memory CD4(+) and CD8(+) T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat. Immunol.* **2**, 2 (2020).
- Zhu, F. C. *et al.* Immunogenicity and safety of a recombinant adenovirus type-5-vectored COVID-19 vaccine in healthy adults aged 18 years or older: A randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet* **396**(10249), 479–488 (2020).
- Jeyanathan, M. *et al.* Immunological considerations for COVID-19 vaccine strategies. *Nat. Rev. Immunol.* **2**, 2 (2020).
- Morens, D. M. & Fauci, A. S. Emerging pandemic diseases: How we got to COVID-19. *Cell* **182**(5), 1077–1092 (2020).

25. Neeffes, J., Jongma, M. L., Paul, P. & Bakke, O. Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nat. Rev. Immunol.* **11**(12), 823–836 (2011).
26. Peters, B., Nielsen, M. & Sette, A. T cell epitope predictions. *Annu. Rev. Immunol.* **38**, 123–145 (2020).
27. Snyder, T. M. *et al.* Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. *MedRxiv*. **2**, 2 (2020).
28. Ritmahan, W., Kesmir, C. & Vroomans, R. M. A. Revealing factors determining immunodominant responses against dominant epitopes. *Immunogenetics* **72**(1–2), 109–118 (2020).
29. Glanville, J. *et al.* Identifying specificity groups in the T cell receptor repertoire. *Nature* **547**(7661), 94–98 (2017).
30. Charini, W. A. *et al.* Clonally diverse CTL response to a dominant viral epitope recognizes potential epitope variants. *J. Immunol.* **167**(9), 4996–5003 (2001).
31. Cornberg, M. *et al.* Narrowed TCR repertoire and viral escape as a consequence of heterologous immunity. *J. Clin. Invest.* **116**(5), 1443–1456 (2006).
32. Luo, W. *et al.* Limited T cell receptor repertoire diversity in tuberculosis patients correlates with clinical severity. *PLoS ONE* **7**(10), e48117 (2012).
33. Ladell, K. *et al.* A molecular basis for the control of preimmune escape variants by HIV-specific CD8+ T cells. *Immunity* **38**(3), 425–436 (2013).
34. Messaoudi, I., Guevara Patino, J. A., Dyllal, R., LeMaoult, J. & Nikolich-Zugich, J. Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. *Science* **298**(5599), 1797–1800 (2002).
35. Peeples, L. News feature: Avoiding pitfalls in the pursuit of a COVID-19 vaccine. *Proc. Natl. Acad. Sci. U.S.A.* **117**(15), 8218–8221 (2020).
36. Altmann, D. M. & Boyton, R. J. SARS-CoV-2 T cell immunity: Specificity, function, durability, and role in protection. *Sci. Immunol.* **5**(49), 2 (2020).
37. Robbiani, D. F. *et al.* Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature* **584**(7821), 437–442 (2020).
38. Noroozi, R. *et al.* Altered cytokine levels and immune responses in patients with SARS-CoV-2 infection and related conditions. *Cytokine* **133**, 155143 (2020).
39. Chen, X. *et al.* Disease severity dictates SARS-CoV-2-specific neutralizing antibody responses in COVID-19. *Signal Transduct. Target Ther.* **5**(1), 180 (2020).
40. Liu, L. *et al.* High neutralizing antibody titer in intensive care unit patients with COVID-19. *Emerg. Microbes Infect.* **9**(1), 1664–1670 (2020).
41. Hoof, I. *et al.* NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* **61**(1), 1–13 (2009).
42. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**(1), 105–132 (1982).
43. Reynisson, B., Alvarez, B., Paul, S., Peters, B. & Nielsen, M. 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 (2020).
44. Team RC. R: A Language and Environment for Statistical Computing. 3.4.4 ed2018.

Author contributions

A.C. and P.C. conceived, designed and organized the work. A.C. wrote the manuscript with support from P.C. V.K., C.M., K.L., A.T., S.H.V. and T.K. did all the assays. S.M., K.B., M.M., A.S., R.S. performed data analysis. K.B. performed single-cell data analysis with input from R.G. P.K. coordinated access to convalescent patient serum. All authors reviewed the manuscript.

Competing interests

All authors are full-time employees of MedGenome Inc. 348 Hatch Drive, Foster City, CA 94404, USA or MedGenome Labs Limited, 258/A 3rd Floor, Narayana Nethralaya, Hosur Road, Bangalore 560099, India and receive salary and benefits.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-021-92521-4>.

Correspondence and requests for materials should be addressed to P.C. or A.C.

Reprints and permissions information is available at www.nature.com/reprints.

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



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

© The Author(s) 2021