



OPEN

An immunodominant NP₁₀₅₋₁₁₃-B*07:02 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease

Yanchun Peng^{1,2,27}, Suet Ling Felce^{2,3,4,27}, Danning Dong^{1,2,5,27}, Frank Penkava^{6,27}, Alexander J. Mentzer^{3,4,27}, Xuan Yao^{2,4,27}, Guihai Liu^{2,4,7,27}, Zixi Yin^{1,2,4,27}, Ji-Li Chen^{1,2,27}, Yongxu Lu⁸, Danielle Wellington^{1,2}, Peter A. C. Wing^{2,4}, Delaney C. C. Dominey-Foy^{2,4}, Chen Jin^{2,4}, Wenbo Wang^{2,4}, Megat Abd Hamid^{2,4}, Ricardo A. Fernandes^{2,4}, Beibei Wang^{2,4}, Anastasia Fries^{3,4}, Xiaodong Zhuang⁴, Neil Ashley⁹, Timothy Rostron¹⁰, Craig Waugh¹¹, Paul Sopp¹¹, Philip Hublitz¹², Ryan Beveridge¹³, Tiong Kit Tan¹, Christina Dold¹⁴, Andrew J. Kwok^{3,4}, Charlotte Rich-Griffin³, Wanwisa Dejnirattisa^{3,4}, Chang Liu^{2,3,4}, Prathiba Kurupati¹, Isar Nassiri^{3,15,16}, Robert A. Watson^{15,16}, Orion Tong^{15,16}, Chelsea A. Taylor^{15,16}, Piyush Kumar Sharma^{15,16}, Bo Sun³, Fabiola Curion^{3,17}, Santiago Revale³, Lucy C. Garner^{16,18}, Kathrin Jansen¹⁹, Ricardo C. Ferreira³, Moustafa Attar¹⁹, Jeremy W. Fry²⁰, Rebecca A. Russell²¹, COMBAT Consortium^{*}, Hans J. Stauss²², William James²³, Alain Townsend^{1,2}, Ling-Pei Ho¹, Paul Klenerman^{18,24}, Juthathip Mongkolsapaya^{2,3,4,25}, Gavin R. Screaton^{2,3,4}, Calliope Dendrou³, Stephen N. Sansom¹⁹, Rachael Bashford-Rogers³, Benny Chain²⁶, Geoffrey L. Smith⁸, Jane A. McKeating^{2,4}, Benjamin P. Fairfax^{15,16}, Paul Bowness⁶, Andrew J. McMichael^{2,4}, Graham Ogg^{1,2}, Julian C. Knight^{2,3,4} ✉ and Tao Dong^{1,2,4} ✉

NP₁₀₅₋₁₁₃-B*07:02-specific CD8⁺ T cell responses are considered among the most dominant in SARS-CoV-2-infected individuals. We found strong association of this response with mild disease. Analysis of NP₁₀₅₋₁₁₃-B*07:02-specific T cell clones and single-cell sequencing were performed concurrently, with functional avidity and antiviral efficacy assessed using an in vitro SARS-CoV-2 infection system, and were correlated with T cell receptor usage, transcriptome signature and disease severity (acute $n = 77$, convalescent $n = 52$). We demonstrated a beneficial association of NP₁₀₅₋₁₁₃-B*07:02-specific T cells in COVID-19 disease progression, linked with expansion of T cell precursors, high functional avidity and antiviral effector function. Broad immune memory pools were narrowed postinfection but NP₁₀₅₋₁₁₃-B*07:02-specific T cells were maintained 6 months after infection with preserved antiviral efficacy to the SARS-CoV-2 Victoria strain, as well as Alpha, Beta, Gamma and Delta variants. Our data show that NP₁₀₅₋₁₁₃-B*07:02-specific T cell responses associate with mild disease and high antiviral efficacy, pointing to inclusion for future vaccine design.

CD8⁺ T cells play a well-documented role in clearing viral infections. Immunodominance is a central feature of CD8⁺ T cell responses in viral infections and understanding the nature of this response for a given infection, where they are shown to be protective, will be critical for the design of vaccines aiming to elicit optimal CD8⁺ T cell responses^{1,2}.

The role of the immunodominant cytotoxic T cell immune response in protection and potential disease pathogenesis of SARS-CoV-2 infection is currently poorly defined. We and others have identified immunodominant T cell epitopes restricted by common human leukocyte antigen (HLA) types³⁻⁶; in particular, we found multiple dominant epitopes in nucleoprotein (NP) restricted by HLA-B*07:02, -B*27:05, -B*40:01, -A*03:01 and -A*11:01. We

also found that multi-functional NP and membrane (M) CD8⁺ T cell responses are associated with mild disease; NP is one of the most common targets for CD8⁺-dominant T cell responses in SARS-CoV-2 infection³.

Among the dominant epitopes identified to date, NP₁₀₅₋₁₁₃-B*07:02 appears to be among the most dominant^{3,4,6,7}; notably, no variants are found within this epitope from over 300,000 sequences in COG-UK global sequence data alignment⁸. This suggests that this epitope would be a good target for inclusion within an improved vaccine design, expanded to stimulate effective CD8⁺ T cell responses, as well as neutralizing antibodies, to protect against newly emergent viral strains that escape antibody responses to spike in some cases⁹.

A full list of affiliations appears at the end of the paper.

Biased *TRBV27* gene usage, with long CDR3 β loops preferentially expressed in NP_{105–113}-B*07:02-specific T cell receptor (TCR), has been observed in both unexposed and COVID-19-recovered individuals¹⁰. The present study suggested a role for cross-reactive responses in COVID-19 based on pre-existing immunity to seasonal coronaviruses or other pathogens. However, a subsequent study suggested that the immunodominant NP_{105–113}-B*07:02 CD8⁺ T cell responses are unlikely to arise from pre-existing cross-reactive memory pools, but rather represent a high frequency of naive T cell precursors found across HLA-B*07:02-expressing individuals⁷.

In this study, we present an in-depth analysis to explore correlations across NP_{105–113}-B*07:02-specific T cell responses, TCR repertoires and disease severity. We saw stronger overall T cell responses in individuals recovered from severe COVID, which may be explained by high exposure to viral protein; however, we found an immunodominant epitope response (HLA-B*07:02 NP_{105–113}-specific CD8⁺) which significantly associated with mild cases. Importantly, this epitope is one of the most dominant CD8⁺ T cell epitopes reported so far by us and others. We examined potential mechanisms of protection using single-cell transcriptome analysis, and functional evaluation of expanded T cell clones bearing the same TCRs as those identified in single-cell analysis. We also assessed the ability of T cell lines and clones to mount effective effector function against cells infected with live SARS-CoV-2 virus and vaccinia virus-expressing SARS-CoV-2 proteins. We found that NP_{105–113}-B*07:02 is the dominant NP response in HLA-B*07:02-positive patients with mild symptoms, with high frequency and higher magnitude when compared with severe cases. Single-cell analysis revealed that preserved beneficial functional phenotypes are associated with protection from severe illness and have better overall antiviral function. In addition, NP_{105–113}-B*07:02-specific T cells can recognize the naturally processed epitope in live virus and recombinant vaccinia virus-infected cells, which correlates with antiviral efficacy.

Results

NP_{105–113}-B*07:02-specific T cell responses are stronger in patients recovered from mild COVID-19 infection. A previous study has identified five dominant CD8⁺ epitopes targeting NP, including the most dominant epitope NP_{105–113} (amino acid sequence SPRWYFYLL) restricted by HLA-B*07:02 (ref. ³). This present study includes 52 individuals who recovered from COVID-19, comprising 30 mild cases and 22 severe cases (including 4 with critical illness; clinical features summarized in Supplementary Table 1 and Extended Data Fig. 1a–c). All the patients were HLA typed and 19 (36.5%) were HLA-B*07:02 positive (10 mild and 9 severe cases; Extended Data Fig. 1d). We proceeded to carry out ex vivo interferon (IFN)- γ ELISpot assays using HLA-B*07:02-positive convalescent samples 1–3 months postinfection. Of HLA-B*07:02 individuals, 79% (15/19) showed responses to this epitope, which accounted for 29% of individuals from the overall cohort (15/52) (Fig. 1a), including 90% (9/10) of individuals recovered from mild and 67% (6/9) from severe disease (Fig. 1b). This further confirms the dominance of this NP_{105–113}-B*07:02 T cell response in our cohort, in particular in individuals recovered from mild illness. In addition, individuals recovered from mild disease made significantly stronger responses to this epitope, compared with those who had recovered from severe disease (Fig. 1c; $P=0.04$). We also observed that this NP_{105–113}-B*07:02-specific response is dominant in mild cases and makes up 60% of overall NP responses of each individual, whereas, in severe cases, the proportion is substantially lower, with an average of 19.5% (Fig. 1d; $P=0.015$). In addition, we did not find HLA-B*07:02 association with disease outcome in our study cohorts (Fig. 1e; 77 acute and 52 convalescent patients). Our data highlight the association of the strength of this dominant epitope-induced T cell response with mild disease outcome and

provide evidence that this link is epitope specific rather than a wider allelic association with HLA-B*07:02.

Strong cytotoxicity and inhibitory receptor expression are associated with disease severity. To explore the mechanisms underlining this association, we sorted NP_{105–113}-B*07:02-specific T cells at a single-cell level with peptide major histocompatibility complex class I (MHC-I) pentamers using flow cytometry. We performed single-cell analysis using SmartSeq2 for peripheral blood mononuclear cell (PBMC) samples from four convalescent patients, including two who recovered from mild COVID-19 infection (C-COV19-005, age 56 years and C-COV19-046, age 76 years) and two who recovered from severe disease in early infection (C-COV19-038, age 44 years and C-COV19-045, age 72 years). TCR sequences and transcriptomic profiles of each single cell were analyzed.

Analysis of single-cell RNA-sequencing (scRNA-seq) data with UMAP visualization and unbiased clustering revealed a homogeneous cell population; therefore we compared gene expression of CD8⁺ NP-specific, sorted single cells isolated from mild ($n=208$ from 2 patients) and severe cases ($n=140$ from 2 patients) by scoring expression levels of manually defined gene sets (Supplementary Table 2). Gene signatures associated with T cell cytotoxicity and inhibitory receptors were analyzed and compared between severity groups. We found that cells from patients who had recovered from severe COVID-19 have significantly higher cytotoxicity gene expression scores (Fig. 2a; $P=0.00032$), with upregulation of *GZMK* ($P=3.02 \times 10^{-3}$) and *GZML* ($P=1.41 \times 10^{-3}$) (encoding granzyme K and granzyme L, respectively) (Fig. 2b). These cells also displayed increased inhibitory receptor expression (Fig. 2c; $P=0.00072$), such as *TIGIT*, *CTLA4* and *HAVCR2* (TIM3). This supports findings published by us and others^{3,11} where patients with severe COVID-19 disease have been exposed to higher antigen loads, and that these cells are still present at 1–3 months convalescence, rather than CD8⁺ central memory T cells.

NP_{105–113}-B*07:02-specific T cells have a highly diverse TCR repertoire. Consistent with findings by other studies^{7,10}, we found that NP_{105–113}-B*07:02-specific T cells from our cohort showed very broad TCR repertoires. Circos plots show paired TCR α and β chains (V and J gene usage) from the four individuals analyzed with SmartSeq2 scRNA-seq (Fig. 3a), and the combined TCR repertoire of all four patients represented by the TCR clonotype (defined separately for each patient combining V gene and CDR3 amino acid sequence) (Fig. 3b). Although the NP_{105–113}-specific TCR repertoire is diverse, with unique pairings of V α and V β genes, we observed that 15/45 (33.3%) of unique V β clonotypes were paired with several distinct V α clonotypes. By contrast, there is only 1/55 (1.8%) V α clonotype that pairs to multiple V β clonotypes; this highlights the importance of studying V β in the TCR repertoire. Further detailed TCR information can be found in Supplementary Table 3.

CDR3 β sequences from patients with mild COVID-19 display higher similarity to naive precursors. Several studies have reported that pre-existing, cross-reactive T cells to SARS-CoV-2 can be detected in unexposed individuals, and these T cells may have resulted from previous human seasonal coronavirus infection^{7,10,12,13}. These studies found TCRs specific to NP_{105–113}-B*07:02 in SARS-CoV-2-unexposed and -infected individuals. These cells were revealed as likely to be naive⁷; this is very different from the central/effector memory phenotype of SARS-CoV-2-specific T cells reported earlier. To investigate this further, we sought to determine what role these T cells might play in the early stages of SARS-CoV-2 infection and COVID-19 disease, and if these cells contribute to the association of mild disease due to their specificity for this NP-dominant epitope.

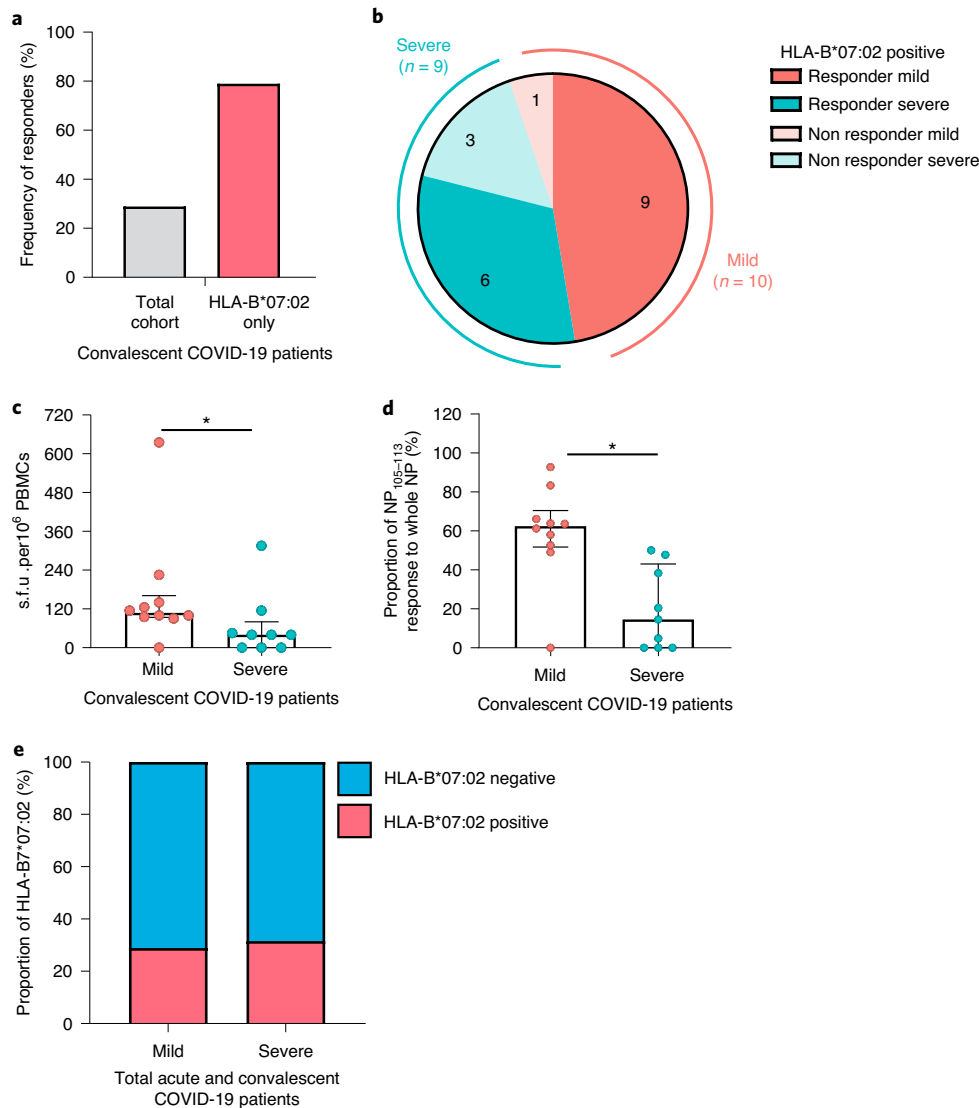


Fig. 1 | Frequency and magnitude of response to the NP₁₀₅₋₁₁₃-B*07:02 epitope in patients with COVID-19. **a**, Frequency of convalescent patients with COVID-19 ($n=52$ total patient cohort, $n=19$ HLA-B*07:02-positive patients only) with T cells responding to NP₁₀₅₋₁₁₃-B*07:02 epitope stimulation. **b**, Frequency of HLA-B*07:02-positive responders ($n=15$) and nonresponders ($n=4$) with mild or severe COVID-19 disease. **c**, Comparison of the magnitude of the response to the NP₁₀₅₋₁₁₃ epitope between HLA-B*07:02-positive convalescent patients with COVID-19 ($n=10$ mild, $n=9$ severe). **d**, Proportion of NP₁₀₅₋₁₁₃-specific response to overall NP response ($n=10$ mild, $n=9$ severe). **e**, Proportion of HLA-B*07:02 individuals compared with combined total acute and convalescent COVID-19 patients ($n=77$ acute, $n=52$ convalescent). Data are presented as medians with interquartile ranges (IQRs) (**c** and **d**). The Mann-Whitney U -test was used for analysis and the two-tailed P value was calculated: $*P < 0.05$. s.f.u., spot-forming units.

To take advantage of the results from our SmartSeq2 scRNA-seq, we first compared TCR sequences from our four convalescent patients with COVID-19 with pre-pandemic TCR sequences from healthy donors, published by Lineburg et al.¹⁰, Nguyen et al.⁷ and another study cohort, COMBAT¹⁴. The COMBAT dataset represents a comprehensive multi-omic blood atlas encompassing acute patients with varying COVID-19 severity (41 mild and 36 severe), and 10 healthy volunteers (pre-pandemic), using bulk TCR sequencing and CITE-Seq, which combines single-cell gene expression and cell-surface protein expression. TCR sequences from the Lineburg and Nguyen datasets have been experimentally validated to be specific for the NP₁₀₅₋₁₁₃ epitope; however, for the COMBAT dataset, we used GLIPH2 analysis¹⁵ to extract TCRs with predicted specificity to this epitope based on convergence with known NP₁₀₅₋₁₁₃-specific TCRs.

We sought to compare NP-specific TCRs from COVID patients and healthy individuals using two different methodologies. First, we

calculated similarity scores for CDR3 β amino acid sequences between pairwise combinations of SmartSeq2 TCRs and pre-pandemic/healthy TCRs. A similarity score of 1 indicates that the pair of CDR3 β sequences are identical, whereas a score of 0 indicates complete dissimilarity. In our convalescent patient cohort, CDR3 β from patients with mild disease are more similar to TCRs from pre-pandemic/healthy individuals, than those from severe patients (Fig. 4a; $P < 2.20 \times 10^{-16}$). Second, we looked at the proportion of TCR sequences from patients with mild and severe disease (acute cases from COMBAT dataset and convalescent cases from previously described SmartSeq2 patients) that can be found in the same convergence groups as sequences from healthy donors, indicating high CDR3 β similarity. Convergence groups containing TCRs from healthy donors appear to contain higher proportions of TCRs from mild cases rather than severe, signifying greater similarity between TCRs from pre-pandemic individuals and patients with mild disease (Fig. 4b; $P < 2.2 \times 10^{-16}$).

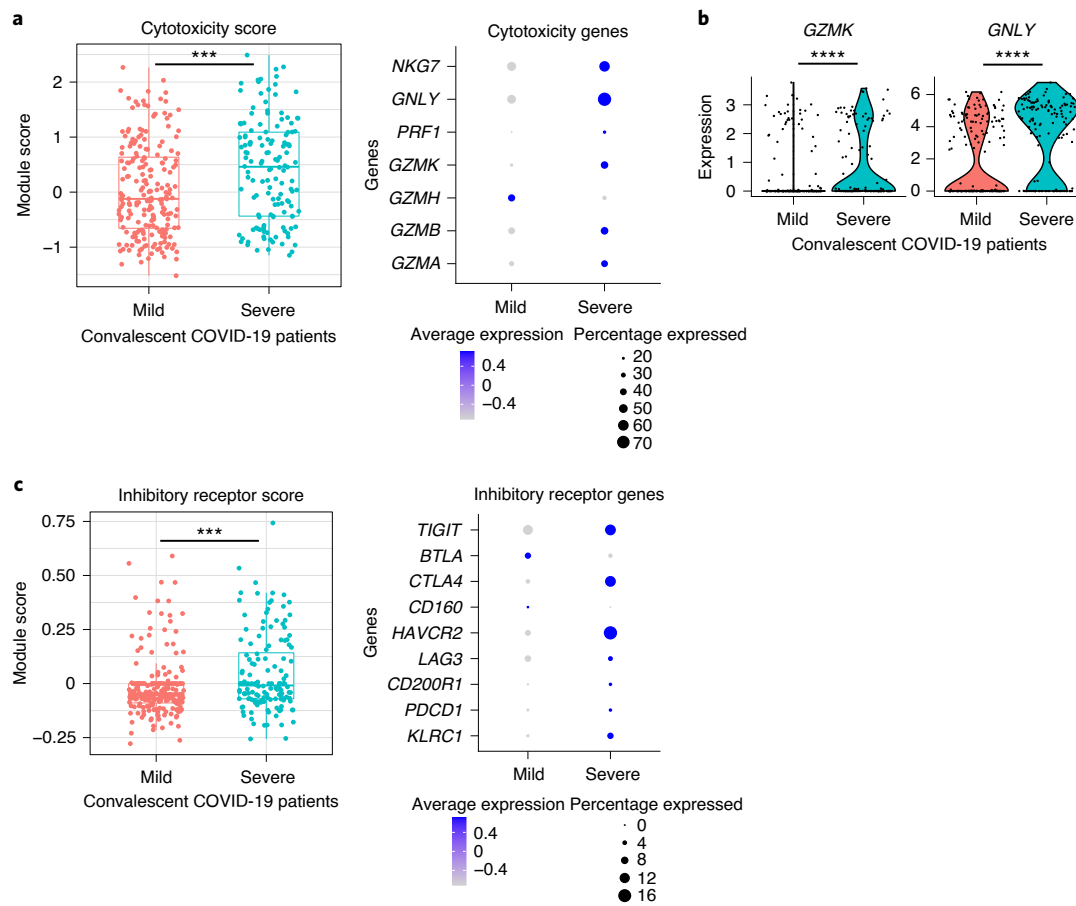


Fig. 2 | Characterization of response to the NP₁₀₅₋₁₁₃-B*07:02 epitope in convalescent HLA-B*07:02-positive patients with mild and severe COVID-19. **a**, Gene sets scored based on single-cell gene expression from a SmartSeq2 RNA-seq dataset comprising two mild and two severe convalescent HLA-B*07:02-positive patients with COVID-19 ($n = 208$ cells from mild cases, $n = 140$ cells from severe cases). Scores for cytotoxic gene expression are shown in the box plot ($P = 0.00032$). Individual genes in the cytotoxic gene set are shown on the right. **b**, Violin plots showing specific expression of cytotoxic genes *GZMK* ($P = 3.02 \times 10^{-5}$) and *GNLY* ($P = 1.41 \times 10^{-9}$). **c**, Box plot showing scores for inhibitory receptor gene set ($n = 208$ cells from mild cases, $n = 140$ cells from severe cases); right: individual genes in gene module. For all box plots, the lower and upper hinges represent the 25–75th percentiles, the central line represents the median, and the whiskers extend to the maximum and minimum values that are no greater than 1.5 \times the IQR. The Mann-Whitney U -test was used for analysis and the two-tailed P value was calculated: *** $P < 0.001$, **** $P < 0.0001$.

We were able to link predicted NP₁₀₅₋₁₁₃-B*07:02TCRs with their corresponding single-cell data from the COMBAT dataset (healthy and acute SARS-CoV-2-infected patients). In this way, we could extract single-cell CITE-seq information from the COMBAT dataset, subsetted specifically to cells with predicted NP₁₀₅₋₁₁₃ specificity. Cellular subtyping of these CD8⁺ NP₁₀₅₋₁₁₃-B*07:02 T cells show a higher proportion of naive T cells in one HLA-B*07:02 healthy individual compared with predominantly T effector memory subtypes in patients with acute COVID-19 ($n = 17$, Fig. 4c). Overall, our data support the report that T cells bearing TCRs specific to NP₁₀₅₋₁₁₃-HLA-B*07:02 in SARS-CoV-2-unexposed individuals are unlikely to have resulted from previous seasonal coronavirus infection⁷. This reinforces the finding that only NP₁₀₅₋₁₁₃-B*07:02-specific T cells from acute HLA-B*07:02-positive patients are exposed to antigen and undergo T cell differentiation, whereas NP₁₀₅₋₁₁₃-specific T cells in pre-pandemic individuals are naive precursors rather than memory cells from a previous crossreactive infection.

Broad range and high functional avidity are associated with clonotype expansion in mild disease. In parallel with single-cell sorting for SmartSeq2, we also sorted, cloned and expanded NP₁₀₅₋₁₁₃-B*07:02-specific T cells from the same convalescent

patients with COVID-19 in vitro^{16,17} to obtain pure clonal T cell populations^{16,17}. We sequenced TCRs from each T cell clone with paired TCR α -chain and β -chain of each clone listed in Supplementary Table 4. When comparing the TCR sequences between T cell clones and ex vivo single cells, in vitro expanded T cell clones are a good representation for the T cells isolated for ex vivo single-cell analysis, with expanded TCRs from ex vivo single cells present as dominant TCRs from the T cell clones (Extended Data Fig. 2a).

To provide a link between T cell clones and single-cell data by their respective TCR sequences, we divided all the T cells, including T cell clones and single cells from SmartSeq2, into 18 groups according to their unique human T cell receptor β variable (*TRBV*) gene usage and CDR3 β sequence (Table 1). T cell functional avidity was measured by IFN- γ ELISpot and calculated from the half-maximal effective concentration (EC_{50}) (Extended Data Fig. 2b and Supplementary Table 5). We found evidence for low and high functional avidity groups (Fig. 5a) based on the EC_{50} of T cell clones, with $EC_{50} < 0.11$ considered to be high-avidity and $EC_{50} > 0.11$ low-avidity T cells. We then aggregated RNA counts from single cells (pseudobulk) to compare differences in gene expression between the two avidity groups. Although there were only seven

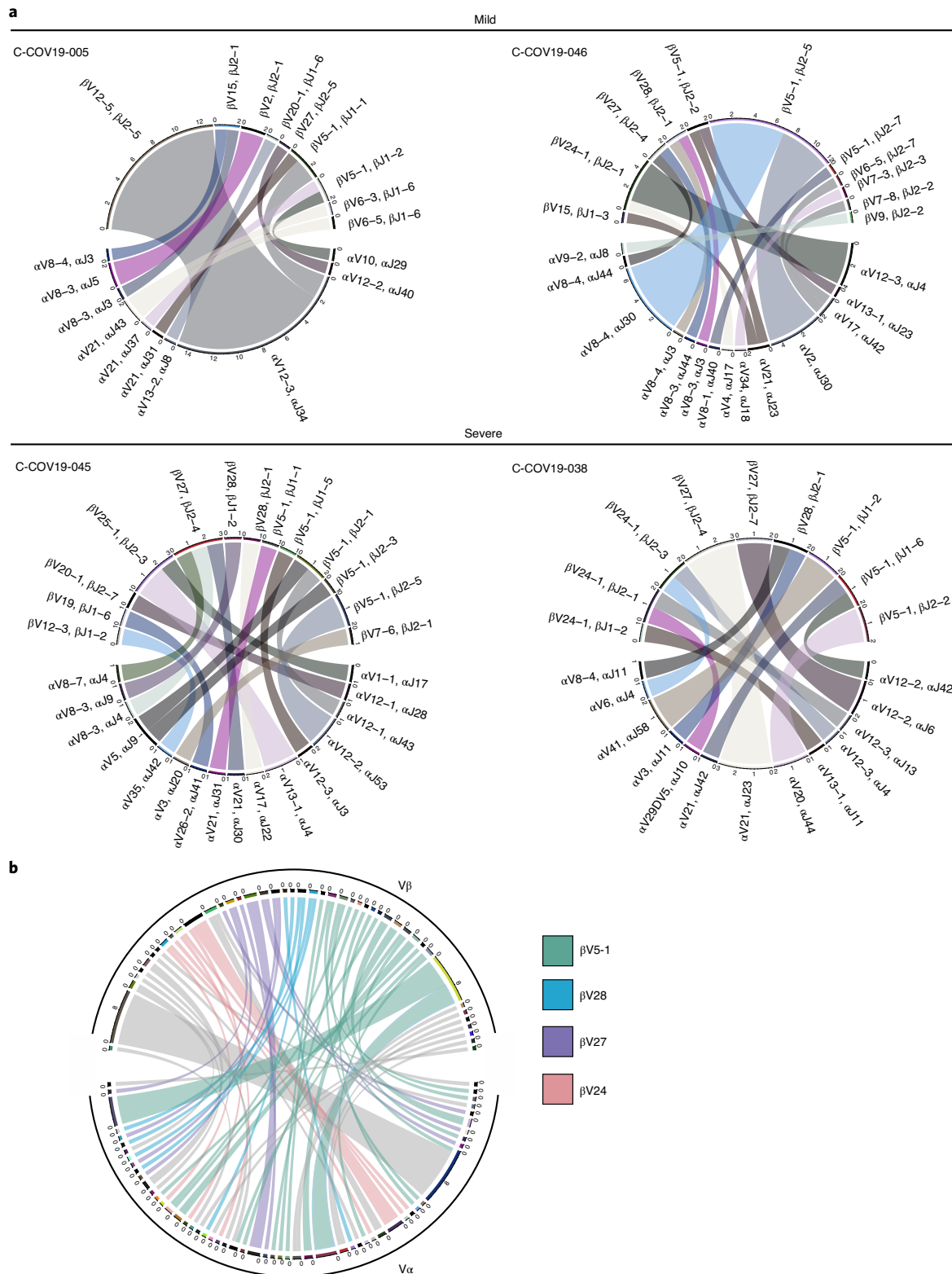


Fig. 3 | Paired $\alpha\beta$ TCR repertoire of NP₁₀₅₋₁₁₃-B*07:02-specific T cells from convalescent patients with COVID-19. **a, Circos plots for each patient depicting $\alpha\beta$ VJ gene usage: two patients with mild disease (C-COV19-005 and C-COV19-046) and two with severe symptoms (C-COV19-045 and C-COV19-038). **b**, Circos plot to show TCR clonotypes for all patients (clonotype defined as patient-specific V-gene usage and CDR3 amino acid sequence for V α and V β). Each line represents a unique clonotype. Clonotypes that have dominant V β gene usage (TRBV5-1, TRBV28, TRBV27 and TRBV24) are highlighted; all others are shown in gray.**

significantly differentially expressed genes (Fig. 5b), possibly as a result of small sample sizes and patient variation, differentially expressed genes of note upregulated in high functional avidity cells

include *IL10RA*, *PARK7* and *LTA4H*. The interaction of interleukin (IL)-10 with IL-10 receptor subunit α (*IL10RA*) expressed on CD8⁺ T cells has been reported to directly decrease CD8⁺ T cell antigen

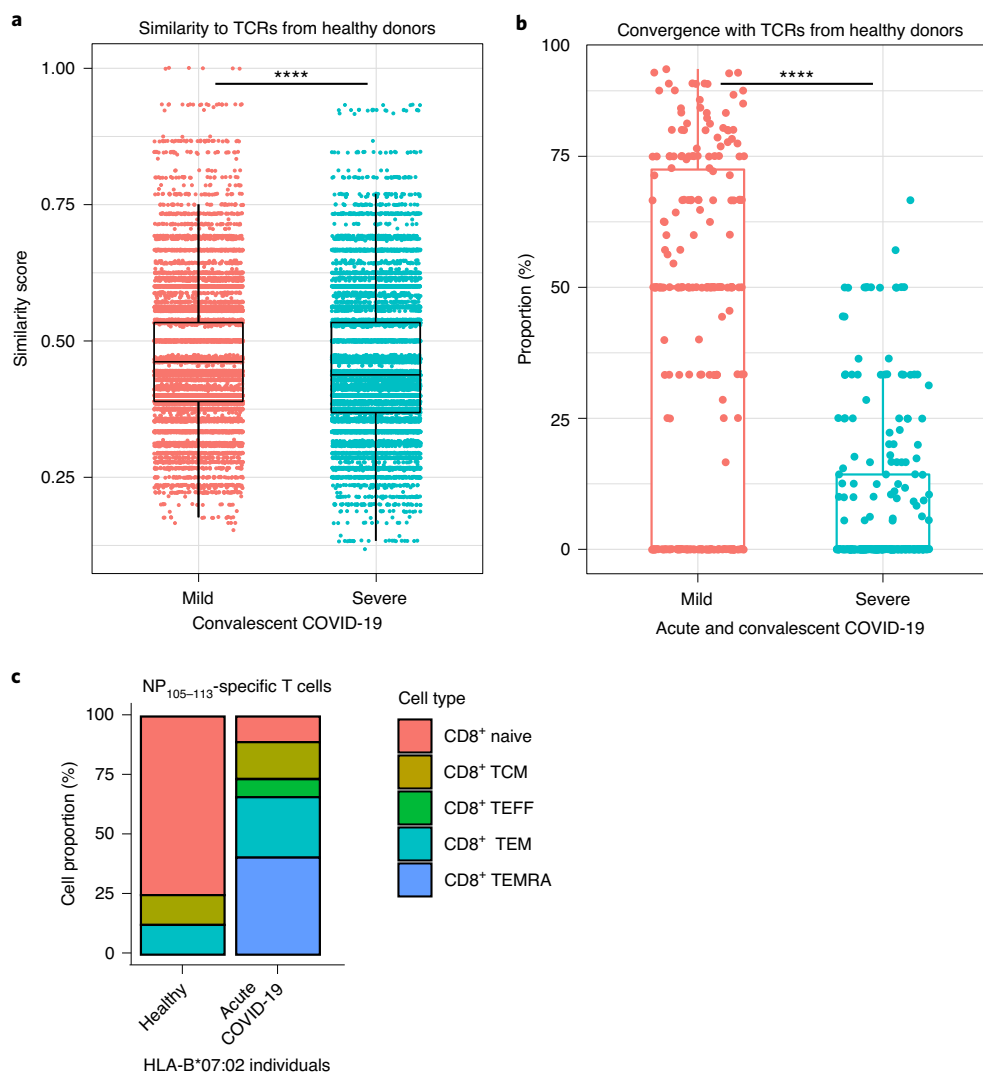


Fig. 4 | Comparison and characterization of NP₁₀₅₋₁₁₃-B*07:02-specific TCRs from acute and convalescent cases of COVID-19. **a**, Similarity scores from pairwise comparisons between TCRs from pre-pandemic individuals (237 TCRs) and 85 unique clonotypes from convalescent patients with COVID-19 (38 mild TCRs versus 47 severe TCRs; $P < 2.20 \times 10^{-16}$). **b**, Proportion of acute and convalescent TCRs from patients with mild and severe COVID-19 found in the same GLIPH2 convergence groups as TCRs from 12 healthy donors (from a total of 738 TCRs from 12 mild patients, 133 TCRs from 7 severe patients and 261 TCRs from healthy individuals in 264 NP₁₀₅₋₁₁₃-B*07:02-predicted convergence groups; mild versus severe: $P < 2.20 \times 10^{-16}$). Each dot on the graph represents a percentage for mild/severe TCRs found in a single convergence group. **c**, Breakdown of CD8⁺ T cell subtypes of T cells with predicted NP₁₀₅₋₁₁₃-B*07:02 specificity from one HLA-B*07:02-positive donor (8 cells) and HLA-B*07:02-positive patients with COVID-19 at acute stage (130 cells from 17 patients with COVID-19). TCM, T central memory; TEFF, T effector; TEM, T effector memory; TEMRA, T effector memory re-expressing CD45RA. For all box plots, the lower and upper hinges represent the 25–75th percentiles, the central line represents the median, and the whiskers extend to maximum and minimum values that are no greater than 1.5x the IQR. The Mann–Whitney *U*-test was used for analysis and the two-tailed *P* value was calculated: **** $P < 0.0001$.

sensitivity in patients with chronic hepatitis C infection¹⁸, whereas Parkinson's disease protein 7 (PARK7) promotes survival and maintains cellular homeostasis in the setting of intracellular stress¹⁹. Leukotriene A4 hydrolase (LTA4H) is an enzyme with known potent anti-inflammatory activity, which functions as an aminopeptidase to degrade a neutrophil chemoattractant Pro-Gly-Pro (PGP) to facilitate the resolution of neutrophilic inflammation and prevent prolonged inflammation with exacerbated pathology and illness²⁰. This supports the idea that high functional avidity T cells undergo stronger antigen stimulation and would therefore start expressing immune-dampening molecules. We further found that patients with mild disease show an increased proportion of high functional avidity TCR clonotypes, which are also more expanded than low functional avidity TCR clonotypes (Fig. 5c), whereas TCR

clonotypes from patients with severe disease show equal expansion between high and low functional avidity TCRs. Therefore, the preferential expansion of high functional avidity TCR clonotypes may contribute to mild disease after SARS-CoV-2 infection.

The strength of T cells responding to naturally processed epitope correlates with their functional avidity. Numerous studies including our own have shown the importance of antigen processing and presentation to T cell recognition of its antigen^{21,22}. Some T cell epitopes may not be processed and presented as efficiently as others, which will subsequently diminish the T cell response to the epitope. To investigate T cell responses to naturally processed and presented viral epitopes, we made vaccinia virus-expressing SARS-CoV-2 viral proteins. We infected autologous

Table 1 | Groups defined by shared TRBV gene usage and CDR3 β sequence between bulk TCR sequencing from T cell clones and single-cell TCR sequencing from ex vivo T cells

Group	CDR3 β	TRBV	TRBJ	Functional avidity
1	CAISEPGTSGGAILDTQYF	TRBV10-3	TRBJ2-3	Low
2	CASGPATSAEQETQYF	TRBV12-5	TRBJ2-5	High
3	CASSILQLGGSNQPQHF	TRBV19	TRBJ1-5	Low
4	CASSVLPGPPRGEQFF	TRBV2	TRBJ2-1	High
5	CSAQVGGNYNSPLHF	TRBV20-1	TRBJ1-6	High
6	CATSDLVTSGDQYF	TRBV24-1	TRBJ2-1	Low
7	CASSGLTSLADTQYF	TRBV25-1	TRBJ2-3	High
8	CASSLITGGAKNIQYF	TRBV27	TRBJ2-4	Low
9	CASSPIAGGRKNIQYF	TRBV27	TRBJ2-4	Low
10	CASSPLTGSARKEQYF	TRBV27	TRBJ2-5	High
11	CASSPLVGERFRKQYF	TRBV27	TRBJ2-5	Low
12	CASSLLAGGFYEQFF	TRBV27	TRBJ2-1	Low
13	CASSPIETAKNIQYF	TRBV28	TRBJ2-4	Low
14	CASSSITTTGAKDGYTF	TRBV28	TRBJ1-2	High
15	CASSLAGAEAFF	TRBV5-1	TRBJ1-1	High
16	CASSLAGGPLHEQFF	TRBV5-1	TRBJ2-1	Low
17	CASSYPGLAPVQETQYF	TRBV5-1	TRBJ2-5	High
18	CASSYLPAGSSYNSPLHF	TRBV6-3	TRBJ1-6	High

Epstein–Barr virus (EBV)-transformed B cell lines (BCLs) with vaccinia virus-expressing NP and cocultured with NP_{105–113}-B*07:02-specific T cell clones. T cell degranulation and cytokine production (CD107a expression and macrophage inflammatory protein (MIP)1 β chemokine production, respectively) were then assessed by intracellular staining after 6h of incubation (Fig. 6a). Gating for CD107a- and/or MIP1 β -producing cells was based on corresponding negative controls (Extended Data Fig. 3a). When compared with the peptide-loaded targets, we found that the response to vaccinia virus-infected BCLs was much weaker, consistent with lower antigen loads. The loading of this naturally processed and presented epitope was equivalent to no more than 3 nM peptide (Extended Data Fig. 3b). Nevertheless, NP vaccinia virus-incubated clones with high CD107a expression showed a negative correlation with their individual EC₅₀ values (Fig. 6b; Spearman's rank correlation coefficient (R) = -0.6176 , $P = 0.0212$), consistent with higher functional avidity resulting in more effective T cell killing. A similar negative correlation was also observed with MIP1 β -producing cells (Fig. 6c; $R = -0.6879$, $P = 0.0082$).

To further investigate the antiviral activity of NP_{105–113}-B*07:02-specific T cells, we established an in vitro SARS-CoV-2 infection system. Briefly, the angiotensin-converting enzyme 2 (ACE2) gene was delivered into autologous EBV-transformed BCLs by lentiviral transduction to enable SARS-CoV-2 infection via ACE2 protein expressed on the cell-surface. ACE2⁺ BCLs were purified by flow sorting and maintained by antibiotic selection, after which cells were subsequently used for SARS-CoV-2 virus infection (Victoria strain). After 48h of incubation, intracellular viral copies were quantified by quantitative (q)PCR, where the reduction of virus replication is calculated as a percentage of virus suppression by T cells (Fig. 6d). We found that the percentage of virus suppression was strongly correlated with their functional avidity (Fig. 6e; $R = -0.7699$, $P = 0.0075$). Therefore, high functional avidity T cells can efficiently inhibit viral replication.

NP_{105–113}-B*07:02-specific T cells are maintained with preserved antiviral efficacy. Six months after infection. To examine whether the memory T cells established postnatural infection could provide sufficient protection against secondary viral infection, we collected PBMCs from three patients (C-COV19-005, C-COV19-045 and C-COV19-046) 6 months after infection and sequenced sorted CD8⁺ NP_{105–113}-B*07:02-specific T cells. We discovered that, 6 months after infection, the TCR repertoire of NP_{105–113}-B*07:02-specific T cells narrows (independent of cell numbers), and the T cell memory pool contains both high and low functional avidity T cells (Fig. 7a). We then isolated and expanded further NP_{105–113}-B*07:02-specific T cell bulk lines from PBMC samples taken 6 months after infection. We assessed the antiviral efficacy of these bulk T cell lines in our in vitro SARS-CoV-2 infection assays. All three T cell lines showed increased MIP1 β and CD107a protein expression after incubation with NP-expressing vaccinia virus (Extended Data Fig. 4), increased tumor necrosis factor (TNF) and CD107a expression after incubation with BCLs infected with SARS-CoV-2 virus (Victoria strain) and current variants of concern (VOCs), including the Delta variant (Fig. 7b and Extended Data Fig. 5). In addition, we found that these antigen-specific bulk cell lines are capable of suppressing SARS-CoV-2 replication (Fig. 7c) and showed strong inhibition against VOCs, including the recently emerged Alpha, Beta, Gamma and Delta SARS-CoV-2 variants (Fig. 7d,e). This is consistent with the evidence of conservation of this NP_{105–113}-B*07:02 epitope, and indicates the protective role of NP_{105–113}-specific T cells in secondary infection against different SARS-CoV-2 variants.

Discussion

Our observation of strong and dominant NP_{105–113}-B*07:02-specific T cell responses in mild cases highlights the possible protective role of this unique and most dominant response found so far in SARS-CoV-2 infection^{3–6}. We found high similarity and convergence of TCRs in HLA-B*07:02-positive healthy and recovered individuals, with naive precursors identified in prepandemic samples supporting previous reports^{7,10}. In addition, T cells from convalescent patients with mild disease show higher functional avidity as well as better effector and antiviral function compared with convalescent patients with severe COVID-19. It is interesting that the immune memory pools postinfection (6 months convalescence) are narrowed but remain proportional; we found no bias toward high or low functional avidity TCRs during immune memory contraction. Moreover, this dominant NP_{105–113}-specific response restricted by HLA-B*07:02 is associated with protection against severe disease, but does not associate with HLA-B*07:02 when analyzed alone.

The highly diverse TCR repertoire of NP_{105–113}-B*07:02-specific T cells in recovered individuals is of particular interest; whether this is a common phenomenon of acute primary virus infection or these responses are unique, with high frequency and broader choice of TCR precursors available, would merit future investigation. The latter is supported by our finding that TCRs in COVID-19-recovered individuals can be similar to those found in prepandemic individuals, in particular patients with mild symptoms. We hypothesize that NP_{105–113}-B*07:02-specific T cell responses play an important role in protecting individuals from severe illness, which is probably due to early priming and expansion of high-frequency naive TCRs specific to this epitope.

We further provide evidence to support our hypothesis by studying a cohort of patients with acute SARS-CoV-2 infection, by analyzing the TCR repertoire in HLA-B*07:02-positive patients. We first found high frequencies of TCR precursors with naive phenotype in HLA-B*07:02-positive healthy donors; this further supports the recent findings, from Nguyen et al.⁷, that these T cell precursors bearing NP-specific TCRs are not due to pre-existing memory from seasonal coronaviruses. We observed that strong cytotoxicity and inhibitory receptor expression are associated with disease severity,

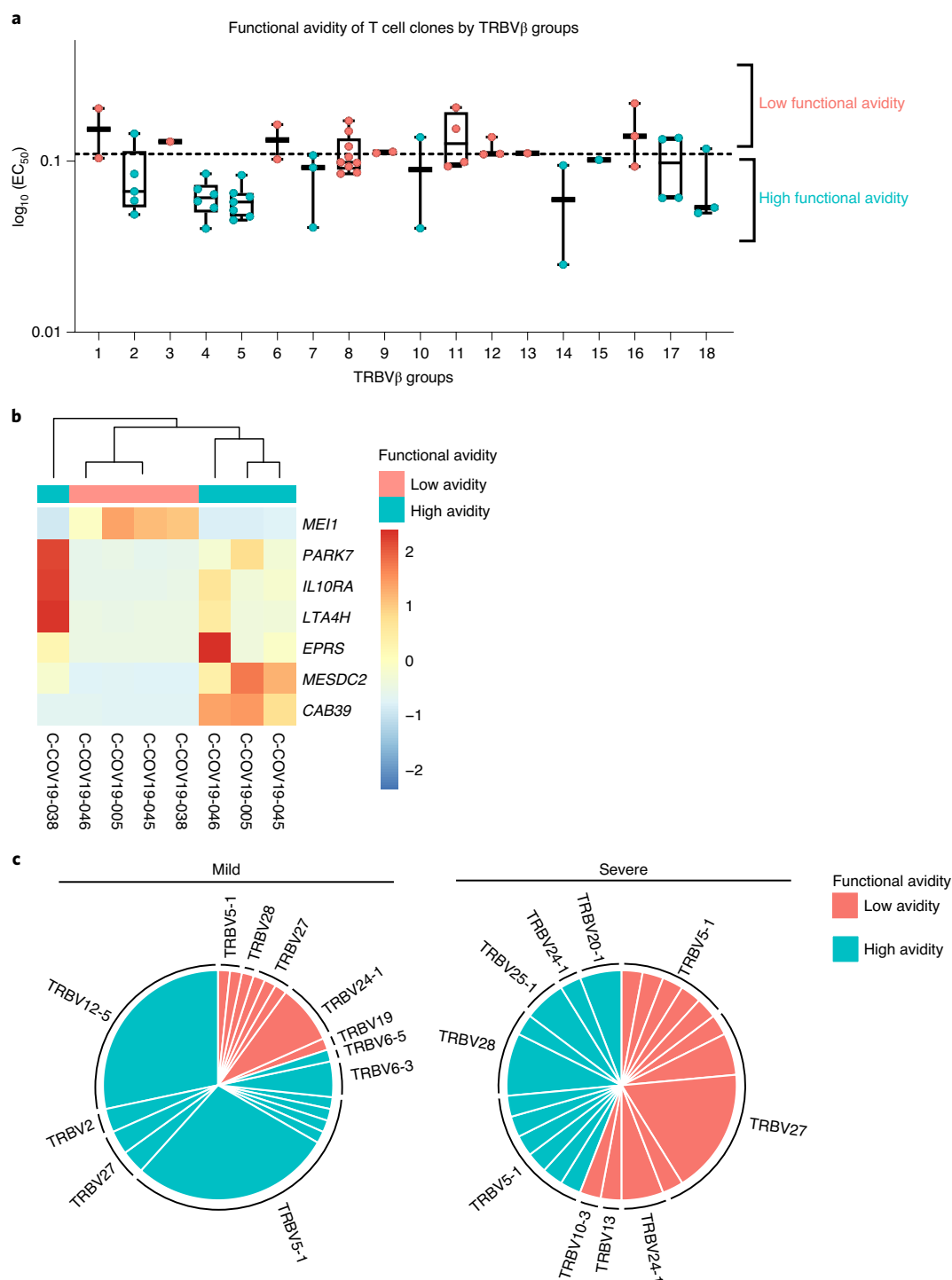


Fig. 5 | Functional avidity and clonotype expansion of NP₁₀₅₋₁₁₃-B*07:02-specific T cells. **a**, Functional avidity of T cell clones by TRBV groups. NP₁₀₅₋₁₁₃-B*07:02-specific T cell clones ($n = 60$) were derived from four convalescent patients with COVID-19 and functional avidity (EC₅₀) was measured by IFN- γ ELISpot assay. On the basis of their TRBV gene usage and CDR3 β sequences, T cell clones and their single-cell counterparts were sorted into 18 distinct groups, and further divided into high or low functional avidity groups (cut-off EC₅₀ = 0.11). The lower and upper hinges of the box on box plots represent the 25–75th percentiles, the central line represents the median, and the whiskers extend to the maximum and minimum values. **b**, Heatmap showing differential gene expression comparing ‘pseudobulk’ high and low functional avidity single cells (88 high-avidity cells and 52 low-avidity cells; the genes shown have adjusted $P < 0.05$). **c**, Comparison of functional avidity and expansion of TCR clonotypes (defined as TRBV gene usage and CDR3 β sequence in each patient) in convalescent patients with mild and severe COVID-19 ($n = 4$).

where NP₁₀₅₋₁₁₃-B*07:02-specific T cells are more activated and well differentiated in individuals recovered from severe illness. This is probably the result of stronger antigen stimulation and expansion during the acute phase of viral infection.

We found overall high functional avidity T cell expansion in mild cases, and that high functional avidity is associated with expression of immune-damping molecules such as IL10RA, PARK7 and LTA4H, which could potentially act to prevent prolonged

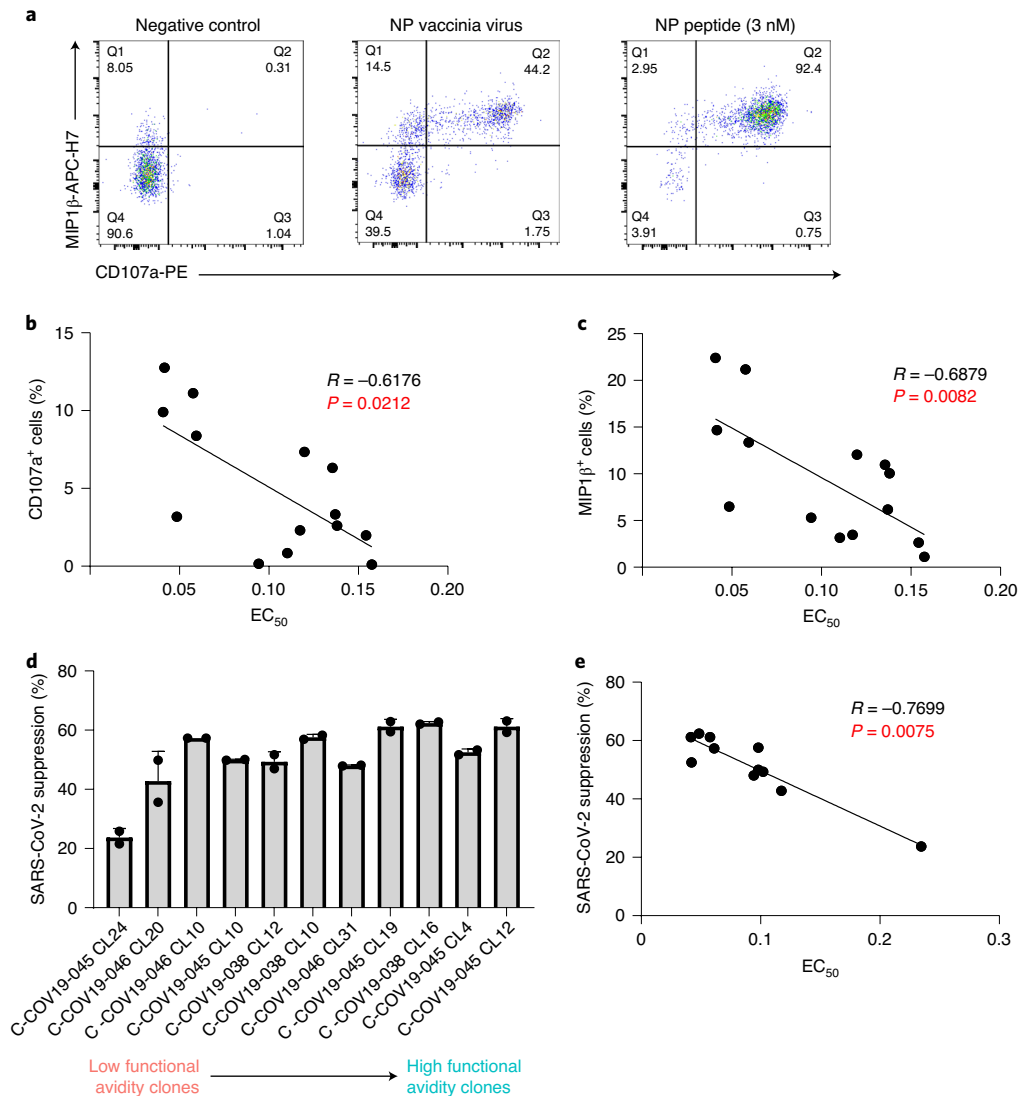


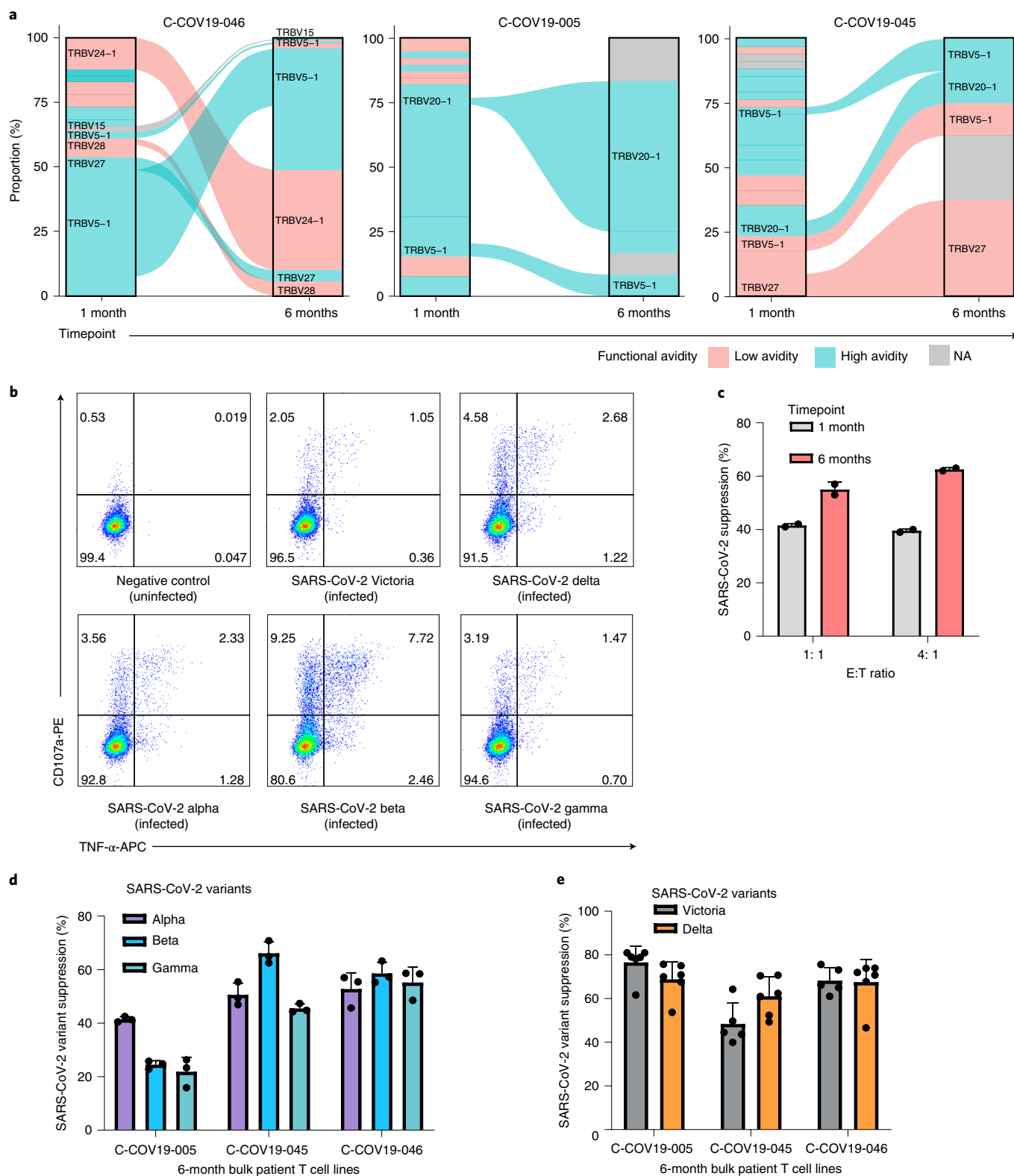
Fig. 6 | Correlation between functional avidity and antiviral efficacy in T cell clones. **a**, Representative ICS flow cytometry plots measuring MIP1 β and CD107a expression on T cell clones incubated with vaccinia virus encoding NP or peptide-loaded (3 nM peptide) antigen-presenting cells. **b**, Correlation plot between CD107a expression on T cell clones incubated with NP-expressing vaccinia virus and their respective EC₅₀ values ($n = 14$, $R = -0.6176$, $P = 0.0212$). **c**, Correlation plot between MIP1 β production of T cell clones incubated with NP-expressing vaccinia virus and their respective EC₅₀ values ($n = 14$, $R = -0.6879$, $P = 0.0082$). **d**, Inhibition of SARS-CoV-2 virus replication (Victoria strain) by T cell clones with different EC₅₀ values and differing functional avidity ($n = 11$). **e**, Correlation plot between percentage of viral suppression by specific T cell clone and its corresponding EC₅₀ value ($n = 11$, $R = -0.7699$, $P = 0.0075$). Spearman's rank correlation coefficient was used for correlation analysis with the two-tailed P value. Bar graph is presented as mean \pm s.d.

inflammation with exacerbated pathology and illness^{18,20,23,24}. In particular, LTA4H has a known function as an aminopeptidase to degrade a neutrophil chemoattractant PGP, facilitating the resolution of neutrophilic inflammation, which is known to be associated with immunopathology in respiratory virus infections such as COVID-19 (ref. ²⁵). This further provides evidence that expansion of high avidity precursors in mild cases contributes to the overall protective immunity from severe illness.

We show that NP₁₀₅₋₁₁₃-B*07:02-specific T cells can respond to cells infected with live SARS-CoV-2 virus as well as emerging viral variants, and most importantly suppress virus replication in infected cells. The magnitude and strength of the response to naturally processed epitopes presented by infected cells correlate with their functional avidity. The proportional expansion with both high and low functional avidity T cells was maintained in CD8⁺ T cell

memory pools after immune memory contraction (at 6 months postinfection), and these cells could suppress virus replication efficiently for all viral variant strains. This is not surprising due to the conservation of this epitope across viral strains, and provides some reassurance that memory T cells generated from natural infection could respond to newly emerged variants and still provide protective immunity.

Taken together, we have demonstrated that, first, we found strong association of NP₁₀₅₋₁₁₃-HLA-B*07:02-specific T cell response with mild disease; second, the protective effect of NP₁₀₅₋₁₁₃-HLA-B*07:02-specific TCRs from severe illness may be due to early expansion of high-frequency naive T cell precursors bearing these TCRs. Moreover, we found that the TCR repertoire is not disturbed after virus infection and immune memory contraction, and that these memory T cells are able to suppress the original SARS-CoV-2



viral strain (contracted by the patient) as well as newly emerged viral strains.

We recognize that there are a number of limitations to the present study, for example, the number of convalescent patients analyzed by single-cell gene expression and TCR sequencing ($n = 4$) is small. Also, the number of NP₁₀₅₋₁₁₃-B*07:02-specific cells from pre-pandemic donors and patients with acute COVID-19 is low, partly

because these cells were not pentamer sorted before analysis. In the present study, we focus on CD8⁺ T cell responses to a single epitope; however, it may be useful in the future to see whether there are any distinct features or features shared with other dominant responses. Although our data support high-frequency naive T cell precursors probably contributing to mild disease outcome, it is also possible, as the consequence of high viral load and overstimulation caused by

Fig. 7 | Characterization of NP₁₀₅₋₁₁₃-B*07:02-specific T cell responses at 6 months convalescence. **a**, TCR repertoires of three patients at 1 month and 6 months convalescence. *TRBV* gene usage of common and expanded TCR clonotypes (defined as *TRBV* and *TRBJ* gene usage) are labeled for clarity. TCR clonotypes colored pink are low functional avidity and blue ones depict high functional avidity; clonotypes colored gray do not have similar TCRs to T cell clones. C-COV19-46 6-month cells were sequenced by 10X single-cell sequencing, and C-COV19-005 and C-COV19-045 by bulk TCR sequencing. NA, not available. **b**, Representative ICS flow cytometry plots measuring TNF- α and CD107a expression on bulk NP₁₀₅₋₁₁₃-specific T cell lines from C-COV19-046 incubated with SARS-CoV-2 Victoria, Alpha, Beta, Gamma or Delta variant-infected BCLs. **c**, Inhibition of SARS-CoV-2 viral replication (Victoria strain) by C-COV19-046 bulk NP₁₀₅₋₁₁₃-specific T cell lines from 1-month (gray bars) and 6-month (red bars) convalescent samples ($n = 2$ biological replicates). Data are shown as mean \pm s.d., representing three independent experiments with similar results. **d**, Antiviral activity of NP₁₀₅₋₁₁₃-specific bulk T cells from 6 months convalescence against SARS-CoV-2 VOCs: Alpha (purple bars), Beta (blue bars) and Gamma (green bars) ($n = 3$ biological replicates). Data are shown as mean \pm s.d., representing three independent experiments with similar results. **e**, Antiviral activity of NP₁₀₅₋₁₁₃-specific bulk T cells from 6 months convalescence against SARS-CoV-2 VOCs: Victoria strain (gray bars) and Delta variant (orange bars) ($n = 6$ biological replicates). Data are shown as mean \pm s.d., representing two independent experiments with similar results.

high functional avidity T cells (with higher proportion of precursor TCRs), leading to exhaustion and depletion during the acute virus infection, which merits further investigation, including larger cohorts sizes. We found that a higher proportion of TCR sequences from mild cases converged with those from prepandemic individuals, although it may be possible that this observation arose from higher numbers of TCRs from mild patients used as input for this convergence analysis. The specifics of antigen loading of this particular epitope, compared with other NP epitopes, as well as variation in levels of protein expression and localization, are also unknown and warrant further investigation.

Online content

Any methods, additional references, Nature Research reporting summaries, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41590-021-01084-z>.

Received: 19 July 2021; Accepted: 26 October 2021;
Published online: 1 December 2021

References

- Yewdell, J. W. Confronting complexity: real-world immunodominance in antiviral CD8⁺ T cell responses. *Immunity* **25**, 533–543 (2006).
- Yewdell, J. W. & Bennink, J. R. Mechanisms of viral interference with MHC class I antigen processing and presentation. *Annu. Rev. Cell Dev. Biol.* **15**, 579–606 (1999).
- 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.* **21**, 1336–1345 (2020).
- Ferretti, A. P. et al. Unbiased screens show CD8⁺ T cells of COVID-19 patients recognize shared epitopes in SARS-CoV-2 that largely reside outside the spike protein. *Immunity* **53**, 1095–1107.e1093 (2020).
- Habel, J. R. et al. Suboptimal SARS-CoV-2-specific CD8⁺ T cell response associated with the prominent HLA-A*02:01 phenotype. *Proc. Natl Acad. Sci. USA* **117**, 24384–24391 (2020).
- Schulien, I. et al. Characterization of pre-existing and induced SARS-CoV-2-specific CD8⁺ T cells. *Nat. Med.* **27**, 78–85 (2021).
- Nguyen, T. H. O. et al. CD8⁺ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope display high naive precursor frequency and TCR promiscuity. *Immunity* **54**, 1066–1082.e1065 (2021).
- de Silva, T. I. et al. The impact of viral mutations on recognition by SARS-CoV-2 specific T-cells. *iScience* **24**, 103353 (2021).
- Zhou, D. et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184**, 2348–2361.e2346 (2021).
- Lineburg, K. E. et al. CD8⁺ T cells specific for an immunodominant SARS-CoV-2 nucleocapsid epitope cross-react with selective seasonal coronaviruses. *Immunity* **54**, 1055–1065.e1055 (2021).
- Sekine, T. et al. Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* **183**, 158–168.e114 (2020).
- 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**, 1489–1501.e1415 (2020).
- Le Bert, N. et al. SARS-CoV-2-specific T cell immunity in cases of COVID-19 and SARS, and uninfected controls. *Nature* **584**, 457–462 (2020).
- Ahern, D. J. et al. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. Preprint at <https://doi.org/10.1101/2021.05.11.21256877> (2021).
- Huang, H., Wang, C., Rubelt, F., Scriba, T. J. & Davis, M. M. Analyzing the *Mycobacterium tuberculosis* immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat. Biotechnol.* **38**, 1194–1202 (2020).
- Abd Hamid, M. et al. Enriched HLA-E and CD94/NKG2A interaction limits antitumor CD8⁺ tumor-infiltrating T lymphocyte responses. *Cancer Immunol. Res.* **7**, 1293–1306 (2019).
- Peng, Y. et al. Boosted influenza-specific T cell responses after H5N1 pandemic live attenuated influenza virus vaccination. *Front. Immunol.* **6**, 287 (2015).
- Smith, L. K. et al. Interleukin-10 directly inhibits CD8⁺ T cell function by enhancing N-glycan branching to decrease antigen sensitivity. *Immunity* **48**, 299–312.e295 (2018).
- Richarme, G. et al. Parkinsonism-associated protein DJ-1/Park7 is a major protein deglycase that repairs methylglyoxal- and glyoxal-glycated cysteine, arginine, and lysine residues. *J. Biol. Chem.* **290**, 1885–1897 (2015).
- Akthar, S. et al. Matrikines are key regulators in modulating the amplitude of lung inflammation in acute pulmonary infection. *Nat. Commun.* **6**, 8423 (2015).
- Rajapaksa, U. S. et al. HLA-B may be more protective against HIV-1 than HLA-A because it resists negative regulatory factor (Nef) mediated down-regulation. *Proc. Natl Acad. Sci. USA* **109**, 13353–13358 (2012).
- Ranasinghe, S. R. et al. The antiviral efficacy of HIV-specific CD8⁺ T-cells to a conserved epitope is heavily dependent on the infecting HIV-1 isolate. *PLoS Pathog.* **7**, e1001341 (2011).
- Singh, Y. et al. Differential effect of DJ-1/PARK7 on development of natural and induced regulatory T cells. *Sci. Rep.* **5**, 17723 (2015).
- Danileviciute, E. et al. PARK7/DJ-1 promotes pyruvate dehydrogenase activity and maintains Treg homeostasis. Preprint at <https://www.biorxiv.org/content/10.1101/2019.12.20.884809v1> (2019).
- Johansson, C. & Kirsebom, F. C. M. Neutrophils in respiratory viral infections. *Mucosal Immunol.* <https://doi.org/10.1038/s41385-021-00397-4> (2021).

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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

¹MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ²Chinese Academy of Medical Science Oxford Institute, University of Oxford, Oxford, UK. ³Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. ⁴Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁵CAMS Key Laboratory of Tumor Immunology and Radiation Therapy, Xinjiang Tumor Hospital, Xinjiang Medical University, Urumqi, China. ⁶Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. ⁷Beijing You'an Hospital, Capital Medical University, Beijing, China. ⁸Department of Pathology, University of Cambridge, Cambridge, UK. ⁹Single Cell Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ¹⁰Sequencing Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ¹¹Flow Cytometry Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ¹²Genome Engineering Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ¹³Virus Screening Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ¹⁴Oxford Vaccine Group, Department of Paediatrics, and NIHR Oxford Biomedical Research Centre, Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, Oxford, UK. ¹⁵Department of Oncology, University of Oxford, Oxford, UK. ¹⁶MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. ¹⁷Helmholtz Center Munich—German Research Center for Environmental Health, Institute of Computational Biology, Neuherberg, Germany. ¹⁸Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ¹⁹Kennedy Institute for Rheumatology, University of Oxford, Oxford, UK. ²⁰ProImmune Limited, Oxford, UK. ²¹Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ²²Institute of Immunity and Transplantation, University College London, London, UK. ²³James & Lillian Martin Centre, Sir William Dunn School of Pathology, University of Oxford, Oxford, UK. ²⁴Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK. ²⁵Dengue Hemorrhagic Fever Research Unit, Office for Research and Development, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand. ²⁶Division of Infection and Immunity, University College London, London, UK. ²⁷These authors contributed equally: Yanchun Peng, Suet Ling Felce, Danning Dong, Frank Penkava, Alexander J. Mentzer, Xuan Yao, Guihai Liu, Zixi Yin, Ji-Li Chen. *A list of members and their affiliations appears in the Supplementary Information. [✉]e-mail: julian.knight@well.ox.ac.uk; tao.dong@imm.ox.ac.uk

Methods

Study participants and clinical definitions. Patients were recruited from the John Radcliffe Hospital in Oxford, UK, between March 2020 and April 2021 by identification of patients hospitalized during the SARS-CoV-2 pandemic. Patients were recruited into the Sepsis Immunomics study and had samples collected during acute disease and convalescence. Patients were sampled at least 28 d after symptom onset. Written informed consent was obtained from all patients. Ethical approval was given by the South Central-Oxford C Research Ethics Committee in England (ref. 19/SC/0296). Clinical definitions were defined as previously described¹.

Generating ACE2-transduced EBV-transformed BCLs. EBV-transformed BCLs were generated as described previously²⁶. The complementary DNA for the human *ACE2* gene (ENSG00000130234) was cloned into a lentiviral vector that allows coexpression of enhanced green fluorescent protein and a puromycin resistance marker (Addgene, plasmid no. 17488). The plasmids were cotransfected with packaging plasmids pMD2.G and psPAX2 into HEK293-TLA using PEIpro (Polyplus). Lentiviral supernatant was collected 48 h and 72 h post-transfection and concentrated by ultracentrifugation. EBV-transformed BCLs were infected by ACE2 lentivirus at a multiplicity of infection (MOI) of 0.1 with $8 \mu\text{g ml}^{-1}$ of polybrene (Sigma-Aldrich) overnight, then washed and cultured for 3–5 d. ACE2-expressing B cells were stained using primary goat anti-human ACE2 antibody (R&D, 1:20) and donkey anti-goat AF647 secondary antibody (Abcam, 1:1,000), followed by cell sorting via flow cytometry. B cells with stable expression of ACE2 were maintained with $0.5 \mu\text{g ml}^{-1}$ of puromycin (Thermo Fisher Scientific). *Mycoplasma* testing was carried out every 4 weeks with all cell lines using MycoAlert detection kit (Lonza).

Generating T cell lines and clones. Short-term SARS-CoV-2-specific T cell lines were established as previously described¹⁷. Briefly, 3×10^6 to 5×10^6 PBMCs were pulsed for 1 h at 37 °C with $10 \mu\text{M}$ peptides, containing T cell epitope regions and cultured in R10 (RPMI 1640 medium with 10% fetal calf serum, 2 mM glutamine and 100 mg ml^{-1} of penicillin–streptomycin) at 2×10^6 cells per well in a 24-well Costar plate. IL-2 was added to a final concentration of 100 U ml^{-1} on day 3 and cultured for a further 10–14 d. T cell clones were generated by sorting HLA-B*07:02 NP_{105–113} pentamer⁺ CD8⁺ T cells at a single-cell level from thawed PBMCs or short-term cell lines. T cell clones were then expanded and maintained as described previously²⁷.

Generating vaccinia virus-expressing SARS-CoV-2 NP. SARS-CoV-2 nucleocapsid (NP) expression vectors (gifts from P. Wang, Shandong University, Shandong, China²⁸) were first digested with KpnI and SacII. The resulting fragment was cloned into vaccinia virus (VACV) expression vector pSC11, which was inserted with a DNA segment encoding KpnI and SacII digestion sites (GGTACCGCGGCCGCCGCGG). The SARS-CoV-2 NP-expressing recombinant VACV (rVACV) was produced as described previously^{29–31}. In brief, HEK293T cells (American Type Culture Collection (ATCC), catalog no. CRL-11268) were transfected with $3 \mu\text{g}$ of pSC11-containing NP with poly(ethylenimine). At 24 h post-transfection, cells were infected with the Lister strain of VACV at an MOI of 1 for 48 h. Infected cells were collected for recombinant virus purification using TK143B cells (ATCC, catalog no. CRL-8303) in $25 \mu\text{g ml}^{-1}$ of bromodeoxyuridine. The NP-expressing rVACV was selected through β -galactosidase staining by supplementing $25 \mu\text{g ml}^{-1}$ of X-gal to an agarose overlay. Master stocks of rVACV were prepared by infection on rabbit RK13 (ATCC, catalog no. CCL37) and titrated on African green monkey BS-C-1 (ATCC, catalog no. CCL26) cells.

SARS-CoV-2 live virus propagation and titration. SARS-CoV-2 Victoria 01/20 strain (BVIC01) and variants of concern—Alpha (Lineage B.1.1.7, 20I/501Y.V1.HMPP1) and Beta (Lineage B.1.351, 20I/501.V2.HV001)—were originally from Public Health England and provided by J. McKeating³². SARS-CoV-2 Gamma (Lineage P.1) was provided by G. Screaton³³. The Delta (B.1.617.2) variant was originally from W. Barclay and T. De Silva (G2P-UK) and provided by G. Screaton³⁴. In brief, Victoria 01/20, Alpha and Beta variants were propagated with Vero E6 cells, whereas SARS-CoV-2 Gamma and Delta variants were propagated with Vero E6/TMPRSS2 (provided by A. Townsend). Naive Vero E6 or Vero E6/TMPRSS2 cells were plated overnight and infected with SARS-CoV-2 at an MOI of 0.003. Cultures were harvested when visible cytopathic effects were observed 48–72 h later. Virus-containing supernatant was aliquoted and stored at $-80 \text{ }^\circ\text{C}$. The viral titer was determined by plaque assay with Vero E6 or Vero E6/TMPRSS2 as previously described, and plaque-forming units per ml were used to calculate the MOI.

IFN- γ ELISpot assay. Ex vivo IFN- γ ELISpot assays were performed using either freshly isolated, cryopreserved PBMCs or antigen-specific T cell clones as described previously³. For ex vivo ELISpots, peptides were added to 2×10^5 PBMCs per test at $2 \mu\text{g ml}^{-1}$ for 16–18 h. When using T cell clones, autologous EBV-transformed BCLs were first loaded with peptides at threefold titrated concentrations and subsequently cocultured with T cells at an effector:target (E:T) ratio of 1:50 for at least 6 h. To quantify antigen-specific responses, data were

collected with AID ELISpot 7.0, mean spots of the control wells were subtracted from the positive wells (phytohemagglutinin stimulation) and the results expressed as spot-forming units (s.f.u.) per 10^6 PBMCs. Responses were considered positive if results were at least three times the mean of the negative control wells and >25 s.f.u. per 10^6 PBMCs. If negative control wells had >30 s.f.u. per 10^6 PBMCs or positive control wells were negative, the results were excluded from further analysis.

Flow cytometric sorting of NP_{105–113}-B*07:02-specific CD8⁺ T cells. NP_{105–113}-B*07:02-specific CD8⁺ T cells were stained with phycoerythrin (PE)-conjugated HLA-B7 NP_{105–113} pentamer (ProImmune). Live/dead fixable Aqua dye (Invitrogen) was used to exclude nonviable cells from the analysis. Cells were washed and stained with the following surface antibodies: CD3-FITC (BD Biosciences), CD8-PerCP-Cy5.5, CD14-BV510, CD19-BV510 and CD16-BV510 (BioLegend). After exclusion of nonviable/CD19⁺/CD14⁺/CD16⁺ cells, CD3⁺CD8⁺pentamer⁺ cells were sorted directly into 96-well PCR plates (Thermo Fisher Scientific) using a BD Fusion sorter or BD FACS Aria III (BD Biosciences) and stored at $-80 \text{ }^\circ\text{C}$ for subsequent analysis.

ScRNA-seq. ScRNA-seq with ex vivo sorted CD8⁺pentamer⁺ T cells was performed using SmartSeq2 (ref. ³⁵) with the following modifications: reverse-transcription (RT) and PCR amplification were performed as described³⁵ with the exception of using ISPCR primer with biotin tagged at the 5' end and increasing the number of cycles to 25. Sequencing libraries were prepared using the Nextera XT Library Preparation Kit (Illumina) and sequencing was performed on Illumina NextSeq sequencing platform with NextSeq Control Software v.4.

Deep sequencing of TCR repertoire of T cell clones. From each T cell clone, 1×10^5 cells were harvested and washed with phosphate-buffered saline (PBS). Total RNA was extracted using RNeasy Plus Micro Kit (QIAGEN), and 100 ng of total RNA from each T cell clone was used to generate full-length TCR repertoire libraries for Illumina Sequencing using a SMARTer Human TCR a/b Profiling Kit (Takara) following the supplier's instructions. The cDNA sequences corresponding to variable regions of TCR- α and/or TCR- β transcripts were amplified with primers, including Illumina indices, allowing for sample barcoding. PCR products were then purified using AMPure beads (Beckman Coulter). The quantity and quality of cDNA libraries were checked on an Agilent 2100 Bioanalyzer system. Sequencing was performed using MiSeq reagent Kit v.3 (600 cycles) on MiSeq (Illumina) with MiSeq Control Software v.2.6.2.1.

Intracellular cytokine staining. Intracellular cytokine staining (ICS) was performed as described previously³. Briefly, T cells were cocultured with peptide-loaded or virus-infected BCLs at an appropriate E:T ratio for a 6 h incubation with GolgiPlug and GolgiStop, and surface stained with PE-anti-CD107a (1:20). Dead cells were labeled using Live/Dead Fixable Aqua dye (Invitrogen); after staining with BV421-anti-CD8 (1:40), cells were then washed, fixed with Cytofix/Cytoperm and stained with AF488-anti-IFN- γ (1:33), APC-anti-TNF- α (eBioscience, 1:500) and APC-H7-anti-MIP1 β (1:33). Negative controls without peptide stimulation or virus infection were run for each sample. All reagents were from BD Bioscience unless otherwise stated. All samples were acquired on Attune NxT Flow Cytometer (software v.3.2.1) and analyzed using FlowJo v.10 software (FlowJo LLC).

Evaluation of T cell response to vaccinia virus infection. EBV-transformed BCLs were infected with Lister strain vaccinia virus at an MOI of 3 for 90–120 min at 37 °C. Cells were washed to remove any virus and incubated overnight in R10 at 37 °C. Cells were counted and cocultured with T cells at an E:T ratio of 1:1. Degranulation (CD107a expression) and cytokine production of T cells were evaluated by ICS as described above.

Evaluation of T cell response to live virus infection. EBV-transformed BCLs expressing ACE2 were infected with SARS-CoV-2 viruses at an MOI of 1 for 120 min at 37 °C. Cells were washed and incubated in R10 at 37 °C. After 24 h, cells were counted and cocultured with T cells at an E:T ratio of 1:1. Degranulation (CD107a expression) and cytokine production of T cells were evaluated by ICS as described above.

Live virus suppression assay. EBV-transformed BCLs expressing ACE2 were infected with SARS-CoV-2 viruses at an MOI of 0.1 for 120 min at 37 °C. Cells were washed and cocultured with T cells at an E:T ratio of 4:1. Control wells containing virus-infected targets without T cells were also included. After 48 h incubation, cells were washed with PBS and lysed with buffer RLT (QIAGEN). RNA was extracted using RNeasy 96 kit (QIAGEN). Virus copies were quantified with Takyon Dry one-step RT-qPCR (Eurogentec) using SARS-CoV-2 (2019-nCoV) CDC qPCR Probe Assay (IDT, ISO 13485:2016) and human β_2 -microglobulin as an endogenous control (Applied Biosystems). The suppression rate was calculated by the percentage reduction of virus replication by T cells.

SmartSeq2 scRNA-seq data processing. BCL files were converted to FASTQ format using bcl2fastq v.2.20.0.422 (Illumina). FASTQ files were aligned to human

genome hg19 using STAR v.2.6.1d³⁶. Reads were counted using featureCounts (subread v.2.0.0 (ref. ³⁷)). The resulting counts matrix was analyzed in R v.4.0.1 using Seurat v.3.9.9.9010 (ref. ³⁸).

SmartSeq2 scRNA-seq analysis. Cells were filtered using the following criteria: minimum number of cells expressing specific gene = 3, minimum number of genes expressed by cell = 200 and maximum number of genes expressed by cell = 4,000. Cells were excluded if they expressed more than 5% mitochondrial genes. Patient-specific cells were integrated using Harmony v.1.0 to remove batch effects. The AddModuleScore function (Seurat) was used to look at the expression of specific gene sets (Supplementary Table 2). The average expression of a gene set was calculated, and the average expression levels of control gene sets were subtracted to generate a score for each cell relating to that particular gene set. Higher scores indicate that that specific signature is expressed more highly in a particular cell compared with the rest of the population. Module scores were plotted using ggplot2 v.3.3.2 (ref. ³⁹).

SmartSeq2 TCR repertoire analysis. TCR sequences were reconstructed from scRNA-seq FASTQ files using MiXCR v.3.0.13 (refs. ^{40,41}) to produce separate TRA and TRB output files for analysis. The output files were parsed into R using tcr v.2.3.2. For paired $\alpha\beta$ TCRs, cells were filtered to retain $1\alpha 1\beta$ or $2\alpha 1\beta$ cells. Circos plots showing paired $\alpha\beta$ TCRs were created using circlize v.0.4.12 (ref. ⁴²). Lists were generated for all 1β cells (regardless of number of α) to use for downstream analysis.

Clustering. Input data for clustering was all 1β from scRNA-seq cells and 1β from bulk sequencing T cell clones. Single cells and clones were grouped by V β usage first; TCRs from either single cells or clones with unique V β gene usage were excluded. Each V β group was broken down into subgroups based on the CDR3 β sequence; any TCRs from either single cells or clones that contained unique CDR3 β sequences were excluded. Only TRBV27, TRBV28 and TRBV5-1 showed multiple CDR3 β sequences with the same gene usage. After plotting EC₅₀ values of T cell clones, groups were classified as low or high functional avidity based on a manually defined cut-off (EC₅₀ = 0.11). This led to a list of 18 groups with unique V β gene usage and CDR3 β sequences shared among the TCRs from single-cell sequencing and bulk T cell clone sequencing.

To group as many single cells into one of these 18 groups, the stringsim function was used (stringdist v.0.9.6 (ref. ⁴³)) to compare the similarity between all SmartSeq2 CDR3 β sequences and each of the 18 CDR3 β from the single cell/clone grouping. A minimum similarity score of 0.7 was used to decide whether a TCR from a single cell should belong to one of the 18 groups. Once allocated, the single cell was annotated as being high or low functional avidity based on its group number.

TCR sequencing from T cell clones (bulk sequencing). BCL files were converted to FASTQ files as described earlier. TCRs were extracted using MiXCR and the resulting output files (TRA and TRB) were parsed into R using tcr as described earlier. TCRs were filtered to retain $1\alpha 1\beta$ for each clone. TCR clonotypes (defined as V β gene usage and CDR3 β sequence) were compared between single TCR and bulk TCR sequencing using ggaluvial v.0.12.2 (ref. ⁴⁴). The predicted functional avidity annotation was overlaid on to the plots using the stringsim function as previously described to classify TCRs into high or low functional avidity groups (minimum score 0.5).

VDJ 10 \times sequencing. Raw BCL files were processed using 10 \times Genomics Cellranger v.5.0.0 (ref. ⁴⁵). For donor deconvolution from multiplexed single-cell data, cellSNP v.0.3.2 (ref. ⁴⁶) was used to generate a list of SNPs from Cellranger output (BAM file). Vireo v.0.5.6 (ref. ⁴⁷) was used to demultiplex the sequencing data into individual patients from the pooled sequenced libraries, based on previously generated SNP-list TCRs from 10 \times sequencing representing 6-month convalescence, and were compared with 1-month convalescence TCRs (SmartSeq2) from the same patient using ggaluvial. The predicted functional avidity annotation was overlaid on to the plots using the stringsim function, as previously described, to classify TCRs into high or low functional avidity groups (minimum score 0.5).

Gene expression analysis and cell subtyping from acute COVID-19 dataset. Normalized single-cell gene expression data for T cells from the COMBAT dataset (level 2 subsets a and b)¹⁴ was annotated with specific T cell subtypes according to COMBAT multimodal analysis, COMBAT TCR chain information and patient metadata. Any cells without both a CD8⁺ multimodal major cell type classification and TCR chain information were excluded from further analysis. A simplified severity grouping based on the World Health Organization's ordinal scale, which ranges from 0 to 8 (https://www.who.int/blueprint/priority-diseases/key-action/COVID-19_Treatment_Trial_Design_Master_Protocol_synopsis_Final_18022020.pdf), was used to classify participants into the following: uninfected (0), mild (1–4), severe (5–7) or death (8).

GLIPH2 analysis. A GLIPH2 CD8⁺ TCR input file was created from the following datasets: COMBAT 10 \times paired-chain single-cell and bulk TCR from all

available participants¹⁴; pentamer-sorted NP_{105–113}-B*07:02-specific TCR sequences and clonally expanded cells used to test functional avidity processed using MiXCR (as described previously); and NP_{105–113}-B*07:02-specific TCR sequences from the Lineburg and Nguyen datasets^{7,10}. Clonotypes were defined as having a unique combination of CDR3 β amino acid sequence, TRBV gene, TRBJ gene and CDR3 α amino acid sequence. Where no or multiple CDR3 α sequences were available for a cell, a not available (NA) value was used for the CDR3 α field in accordance with GLIPH2 input guidelines. For each clonotype, additional information indicating dataset origin was appended as part of the 'condition' field. For the 10 \times COMBAT dataset, CD8⁺ clonotypes were distinguished from CD4⁺ clonotypes based on the multimodal classification of cells within each clone.

A matching GLIPH2 participant HLA input file was created using COMBAT formal HLA-typing data and, where no formal typing was available, from imputed HLA typing^{34,4}, in addition to published HLA data relating to the Lineburg and Nguyen datasets^{7,10}.

The GLIPH2 irtools.centos v.0.01 (ref. ¹⁵) was run on a CentOS Linux platform (release 8/3/2011) using the CD8⁺ TCR and HLA input files above, together with CD8⁺-specific V-gene usage, CDR3 length and TCR reference files from the GLIPH2 repository and using the following parameters: local_min_pvalue = 0.001; p_depth = 1000; global_convergence_cutoff = 1; simulation_depth = 1000; kmer_min_depth = 3; local_min_OVE = 10; algorithm = GLIPH2; all_aa_interchangeable = 1; number_of_hla_field = 1; and hla_association_cutoff = 0.050000. A GLIPH score summary file was then programmatically curated, identifying convergence groups containing TCRs known to be NP_{105–113}-B*07:02 specific as described previously, with associated GLIPH2 scoring and HLA prediction.

Convergence groups from this file were further categorized as being associated with or lacking association with HLA-B*07:02 based on having a GLIPH2 HLA score <0.05 or \geq 0.05, respectively. Only clonotypes belonging to a HLA-B*07-associated convergence group, which were from participants known to have a HLA-B*07:02 allele, were deemed to be HLA-B*07:02-positive TCRs. Any clonotypes from convergence groups lacking HLA-B*07:02 association, but belonging to patients with a HLA-B*07:02 allele, were deemed ambiguous and excluded from the HLA-B*07:02-negative clonotype set.

Similarity between pre-pandemic and convalescent COVID-19 TCRs. NP_{105–113}-specific TCRs from pre-pandemic individuals (predicted from the COMBAT dataset or experimentally defined by the Lineburg and Nguyen datasets^{7,10}) were compiled to form a single list of sequences (237 TCRs). Similarity scores were calculated from pairwise comparisons between each CDR3 β sequence from the pre-pandemic/healthy list and each CDR3 β sequence from 85 unique clonotypes of 4 convalescent patients with COVID-19 (clonotype defined per patient, TRBV gene usage and CDR3 β sequence). A score of 1 indicates total similarity whereas a score of 0 is total dissimilarity. Each score was plotted on a box plot using ggplot2.

Pseudobulk and differential gene expression. RNA counts from SmartSeq2 single cells were aggregated into groups based on patient origin and high/low functional avidity, and converted to a Single Cell Experiment (v.1.10.1) object⁴⁸. Differential gene expression was conducted using DESeq2 v.1.28.1 on aggregated (pseudobulk) counts. Significant genes were visualized on a heatmap using pheatmap v.1.0.12.

Statistics. A Mann–Whitney nonparametric *U*-test was used to compare two groups (*R*); other statistical tests were carried out using GraphPad Prism. Nonlinear regression with variable slope (four parameters) in a dose–response–stimulation model was used for calculating the EC₅₀ of T cell clones. Spearman's rank correlation coefficient was used for correlation analysis. NS, not significant; **P* < 0.05, ***P* < 0.01, ****P* < 0.001, *****P* < 0.0001.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The raw data from all the main and supplementary figures are available on request. In addition the following published datasets were used: Lineburg et al.¹⁰ (<https://doi.org/10.1016/j.immuni.2021.04.006>), Nguyen et al.⁷ (<https://doi.org/10.1016/j.immuni.2021.04.009>) and COMBAT (<https://doi.org/10.1101/2021.05.11.21256877>).

References

- Dong, T. et al. HIV-specific cytotoxic T cells from long-term survivors select a unique T cell receptor. *J. Exp. Med.* **200**, 1547–1557 (2004).
- Abd Hamid, M. et al. Self-maintaining CD103⁺ cancer-specific T cells are highly energetic with rapid cytotoxic and effector responses. *Cancer Immunol. Res.* **8**, 203–216 (2020).
- Zhang, J. et al. A systemic and molecular study of subcellular localization of SARS-CoV-2 proteins. *Signal Transduct. Target. Ther.* **5**, 269 (2020).
- Mackett, M., Smith, G. L. & Moss, B. General method for production and selection of infectious vaccinia virus recombinants expressing foreign genes. *J. Virol.* **49**, 857–864 (1984).

30. Smith, G. L., Murphy, B. R. & Moss, B. Construction and characterization of an infectious vaccinia virus recombinant that expresses the influenza hemagglutinin gene and induces resistance to influenza virus infection in hamsters. *Proc. Natl Acad. Sci. USA* **80**, 7155–7159 (1983).
31. Chakrabarti, S., Brechling, K. & Moss, B. Vaccinia virus expression vector: coexpression of beta-galactosidase provides visual screening of recombinant virus plaques. *Mol. Cell Biol.* **5**, 3403–3409 (1985).
32. Wing, P. A. C. et al. Hypoxic and pharmacological activation of HIF inhibits SARS-CoV-2 infection of lung epithelial cells. *Cell Rep.* **35**, 109020 (2021).
33. Dejnirattisai, W. et al. Antibody evasion by the P.1 strain of SARS-CoV-2. *Cell* **184**, 2939–2954 e2939 (2021).
34. Liu, C. et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **184**, 4220–4236.e4213 (2021).
35. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
36. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
37. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
38. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 e1821 (2019).
39. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis. Use R!*, 2nd edn, 1 (Springer International Publishing, 2016).
40. Bolotin, D. A. et al. Antigen receptor repertoire profiling from RNA-seq data. *Nat. Biotechnol.* **35**, 908–911 (2017).
41. Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
42. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812 (2014).
43. van der Loo, M. The stringdist package for approximate string matching. *R J.* **6**, 111–122 (2014).
44. Brunson, J. C. ggalluvial: layered grammar for alluvial plots. *J. Open Source Software* <https://doi.org/10.21105/joss.02017> (2020).
45. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
46. Huang, X. & Huang, Y. Cellsnp-lite: an efficient tool for genotyping single cells. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btab358> (2021).
47. Huang, Y., McCarthy, D. J. & Stegle, O. Vireo: Bayesian demultiplexing of pooled single-cell RNA-seq data without genotype reference. *Genome Biol.* **20**, 273 (2019).
48. Amezquita, R. A. et al. Orchestrating single-cell analysis with Bioconductor. *Nat. Methods* **17**, 137–145 (2020).

Acknowledgements

We thank all the participants for donating their samples and data for these analyses, and the research teams involved in the consenting, recruitment and sampling of these participants. We thank K. Clark and S.-A. Clark from WIMM flow cytometry facility for their help with cell sorting. The B.1.617.2 (Delta) variant was provided by W. Barclay and T. De Silva from the UK RI-funded Genotype to Phenotype Consortium (G2P-UK). This work is supported by the UK Medical Research Council (MRC, to T.D., G.O., Y.P. and J.C.); Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS), China (grant no. 2018-I2M-2-002 to T.D., Y.P., S.L.F., X.Y., G.L., D.D., J.C.K., G.O., G.S., J.M., A.M., J.A.M. and A.T.); National Institutes of Health, National

Key R&D Program of China (no. 2020YFE0202400 to T.D.); China Scholarship Council (to Z.Y., G.L., C.J. and C.L.). The present study is also funded by the NIHR Oxford Biomedical Research Centre (L.P.H., G.O., J.C.K., P.B. and A.J.M.), Senior Investigator Award (to G.O.) and Clinical Research Network (to G.O.); Schmidt Futures (G.R.S.); the National Institute for Health Research (NIHR) (UKRIDHSC COVID-19 Rapid Response Rolling Call, grant no. COV19-RECLAS), WT109965MA, Huo Family Foundation, UK-CIC and DHSC (PITCH; to P.K.). J.C.K. is supported by a Wellcome Investigator award (no. 204969/Z/16/Z). The McKeating laboratory is funded by a Wellcome Investigator Award (to J.A.M., no. 200838/Z/16/Z) and UK MRC project grant (no. MR/R022011/1). Initial funding for the VSF was provided by the Oxford Biomedical Research Centre and Cancer Research UK. We thank the donors to the University of Oxford's COVID-19 Research Response Fund for the philanthropic support of the COMBAT study. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the UK Department of Health. This work uses data provided by patients and collected by the NHS as part of their care and support #DataSavesLives.

Author contributions

T.D. and J.C.K. conceptualized the project. T.D. and Y.P. designed and supervised the T cell experiments. Y.P., D.D., X.Y., G.L., Z.Y. and J.C. performed all the T cell experiments. Y.L. and G.S. provided vaccinia virus. P.A.C.W. and X.Z. assisted with virus infection. T.R. performed HLA typing and next-generation sequencing. N.A. prepared cDNA libraries for sequencing. P.H., R.B. and T.K.T. made the ACE2 constructs and lentivirus. J.W.F. provided MHC-I pentamers. J.C.K., A.J.M., L.P.H. and A.F. established clinical cohorts and collected clinical samples and data. D.W., D.D.-F., C.J., W.W., M.A.H., B.W., C.D. and W.D. processed clinical samples. R.A.F., C.W., P.S., A.K., C.R.-G., R.B.-R., C.L., P.K., I.N., R.A.W., O.T., C.A.T., P.K.S., B.S., F.C., S.R., L.C.G., K.J., R.C.F., M.A., R.A.R., C.D., S.N.S., B.F., J.A.M., W.J., A.T., G.L.S. and J.M. provided technical assistance and critical reagents. Y.P., S.L.F., F.P. and G.L. analyzed the data. T.D., J.C.K. and Y.P. supervised the data analysis. T.D., Y.P. and S.L.F. wrote the original draft of the manuscript. J.C.K., G.O., A.M., G.L.S., H.J.S., P.K., B.C., P.B. and S.L.F. reviewed and edited the manuscript and figures.

Competing interests

H.J.S. is a consultant and shareholder at Quell Therapeutics, a consultant at PanCancerT and a recipient of research funds from Apollo Therapeutics. The remaining authors declare no competing interests.

Additional information

Extended data are available for this paper at <https://doi.org/10.1038/s41590-021-01084-z>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41590-021-01084-z>.

Correspondence and requests for materials should be addressed to Julian C. Knight or Tao Dong.

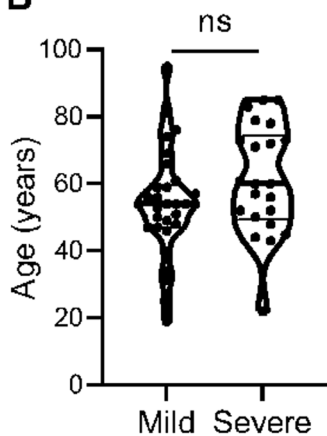
Peer review information *Nature Immunology* thanks Robert Thimme and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer review reports are available. Zoltan Fehervari was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

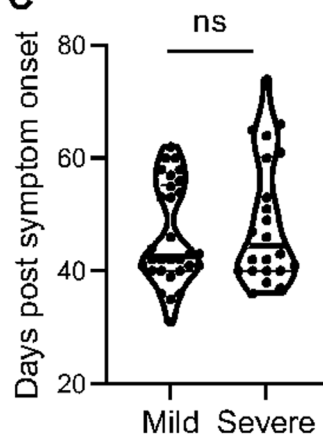
A

	Mild Disease (N=30, 1 asymptomatic)	Severe Disease (N=22, 4 critical)
Age, y, median (IQR)	54.0(47.0-59.5)	60.0(49.5-74.3)
Male sex	20(66.7)	14(63.6)
Days post symptom, median (IQR)	42.5(40.0-55.3)	44.5(40.0-60.3)

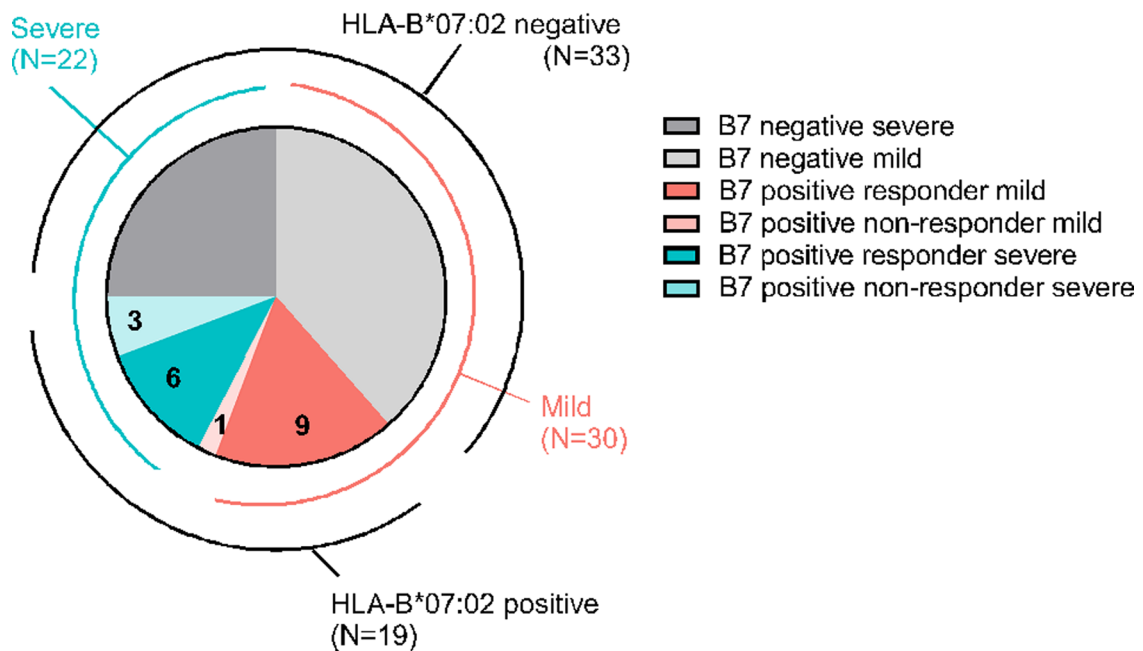
B



C

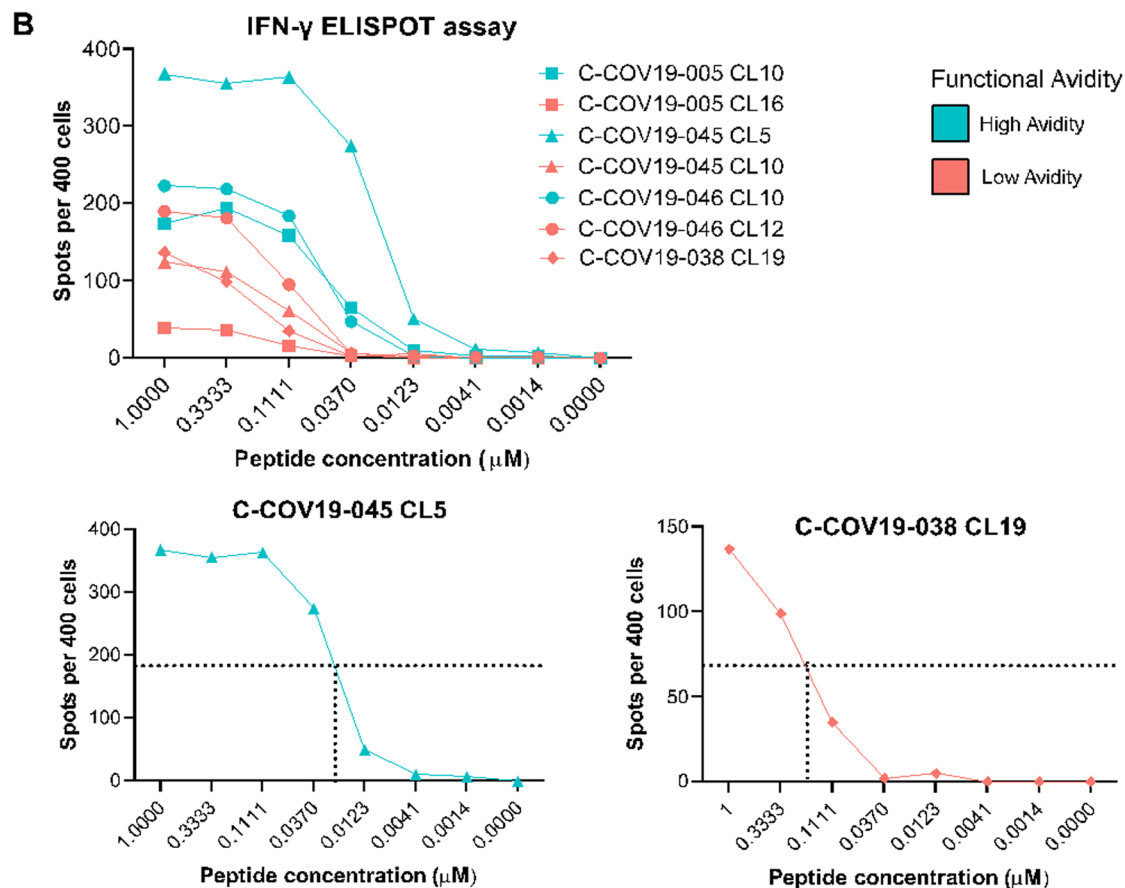
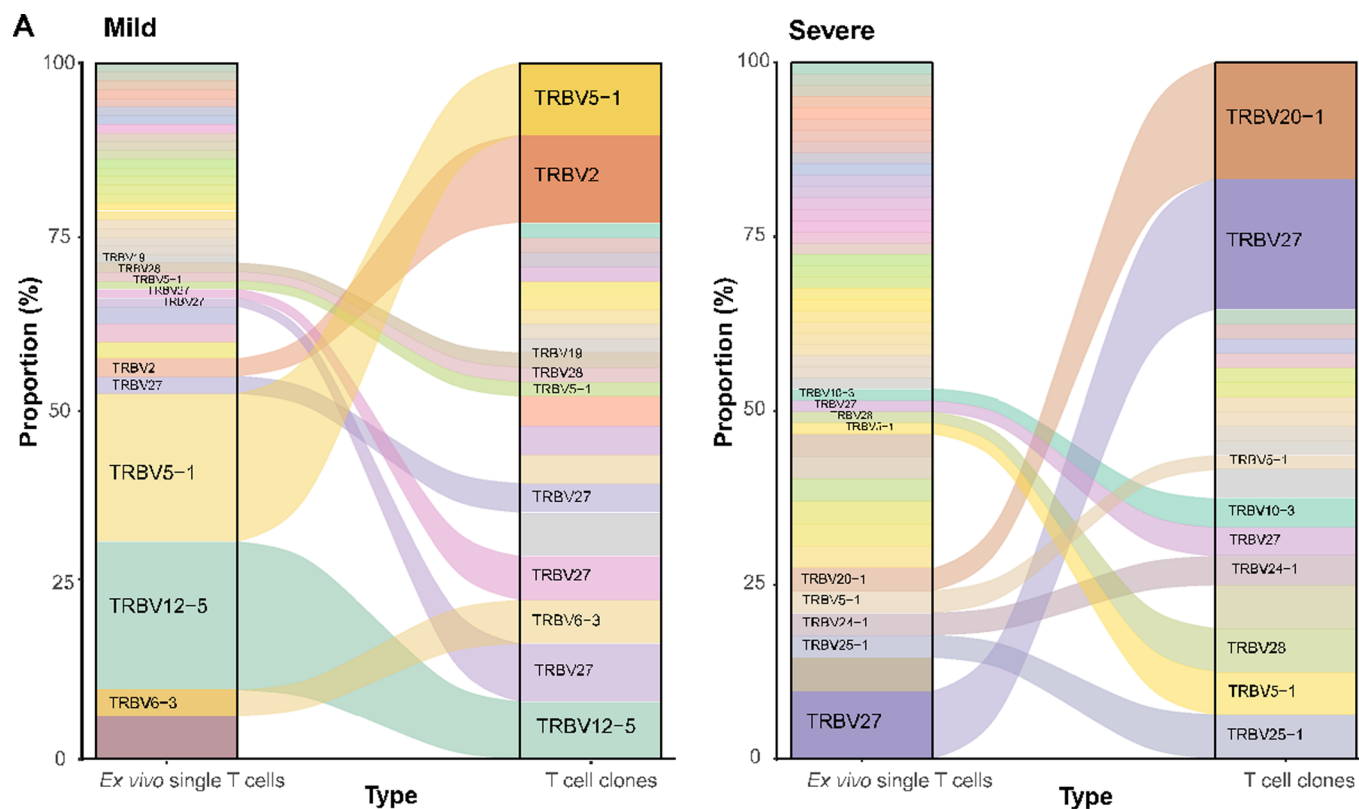


D



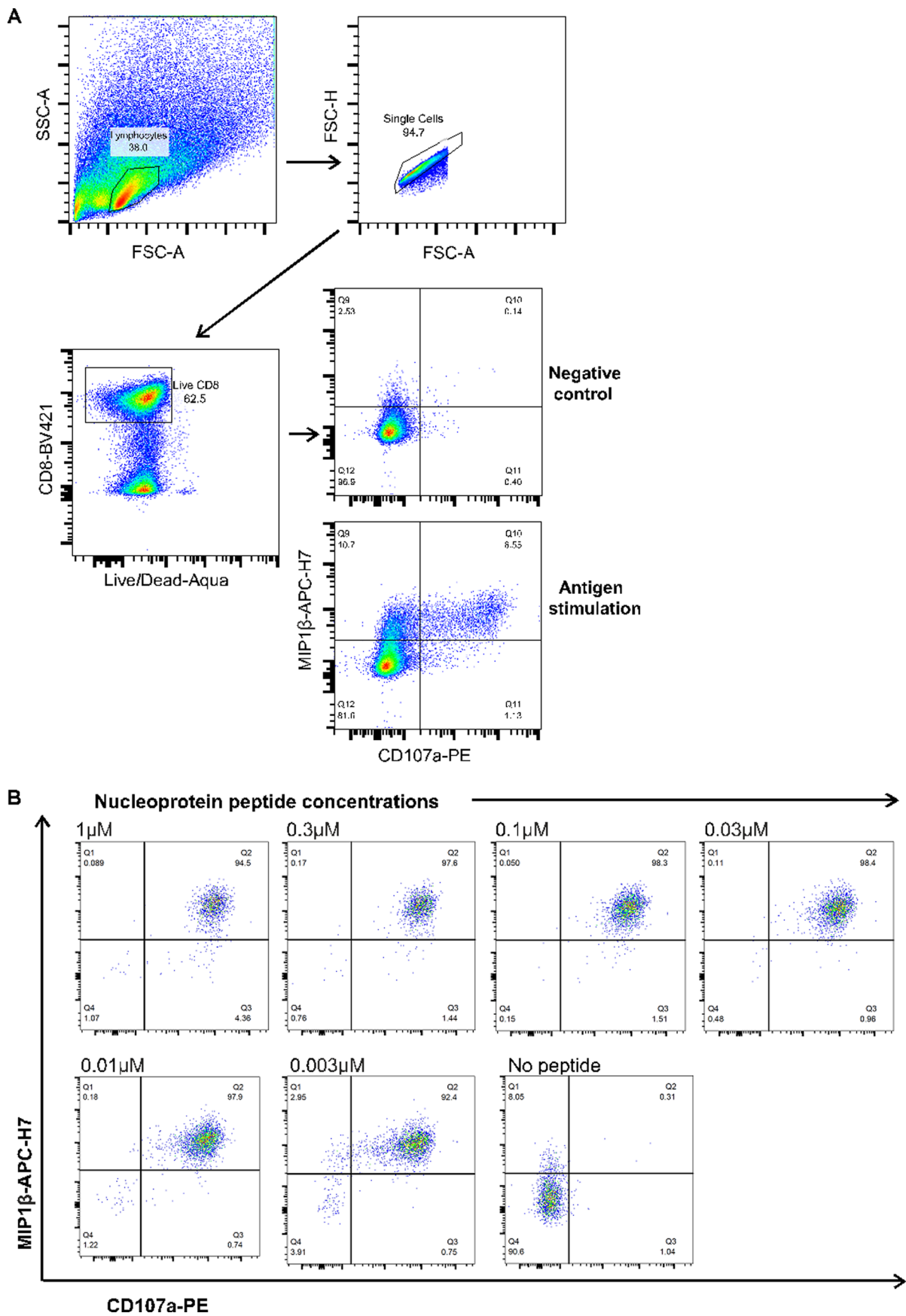
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Patient characteristics. (a) Distribution of age, gender and days post symptom onset when sampling of SARS-CoV-2 infected patients. (b) and (c) Comparison of age (ns, $P=0.116$) and days post symptom (ns, $P=0.5711$) when sampling between the patient groups with mild and severe symptoms. Unpaired t-test with Welch's correction and Mann-Whitney test were used for data analysis of (B) and (C) respectively. Two tailed P value was calculated. (d) Breakdown of patient numbers and severity; HLA-B*07:02 negative (mild N=20 light grey, N=13 dark grey); HLA-B*07:02 positive responders (mild N=9 red, severe N=6 blue); HLA-B*07:02 positive non-responders (mild N=1 light red, severe N=3 light blue).

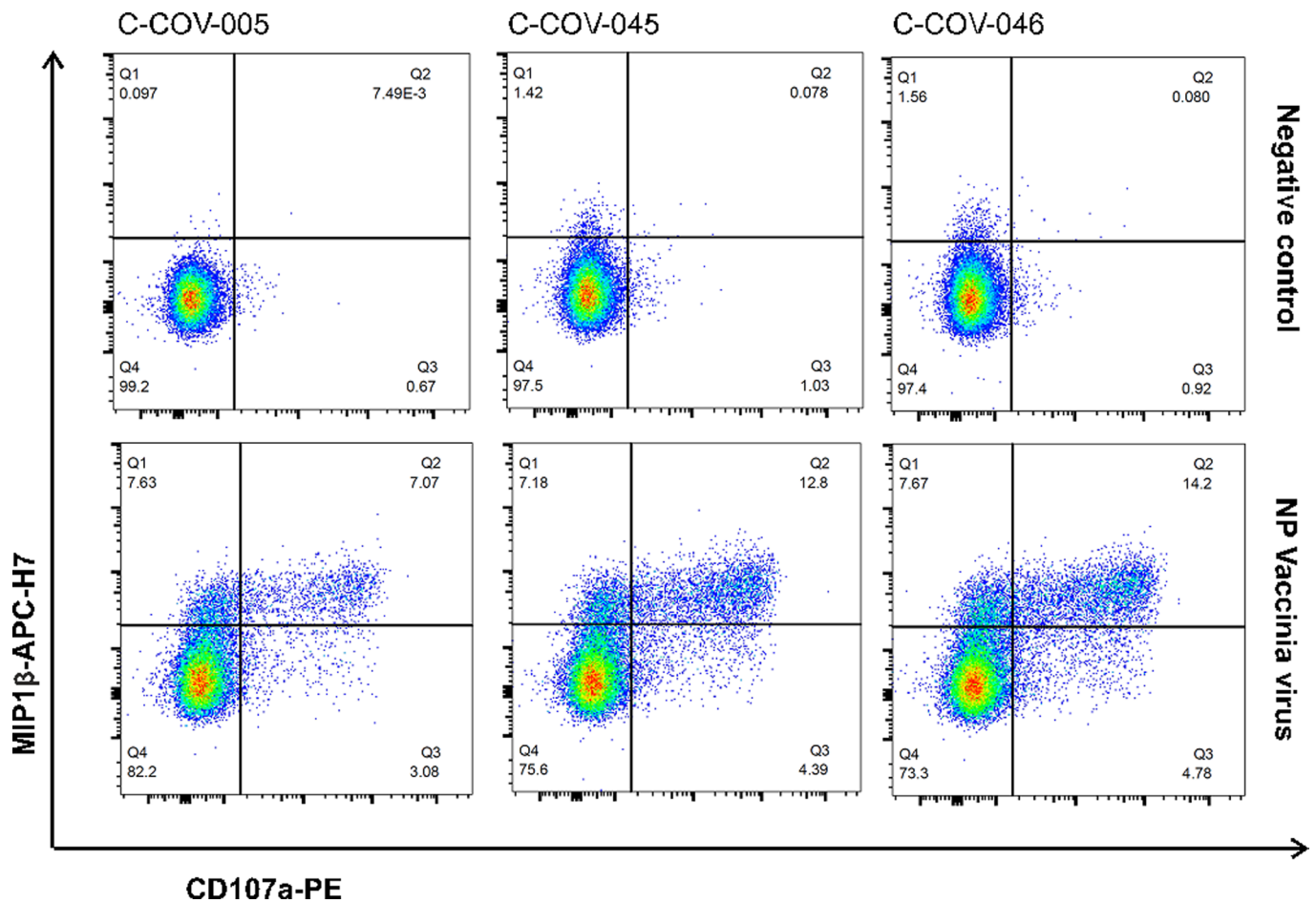


Extended Data Fig. 2 | See next page for caption.

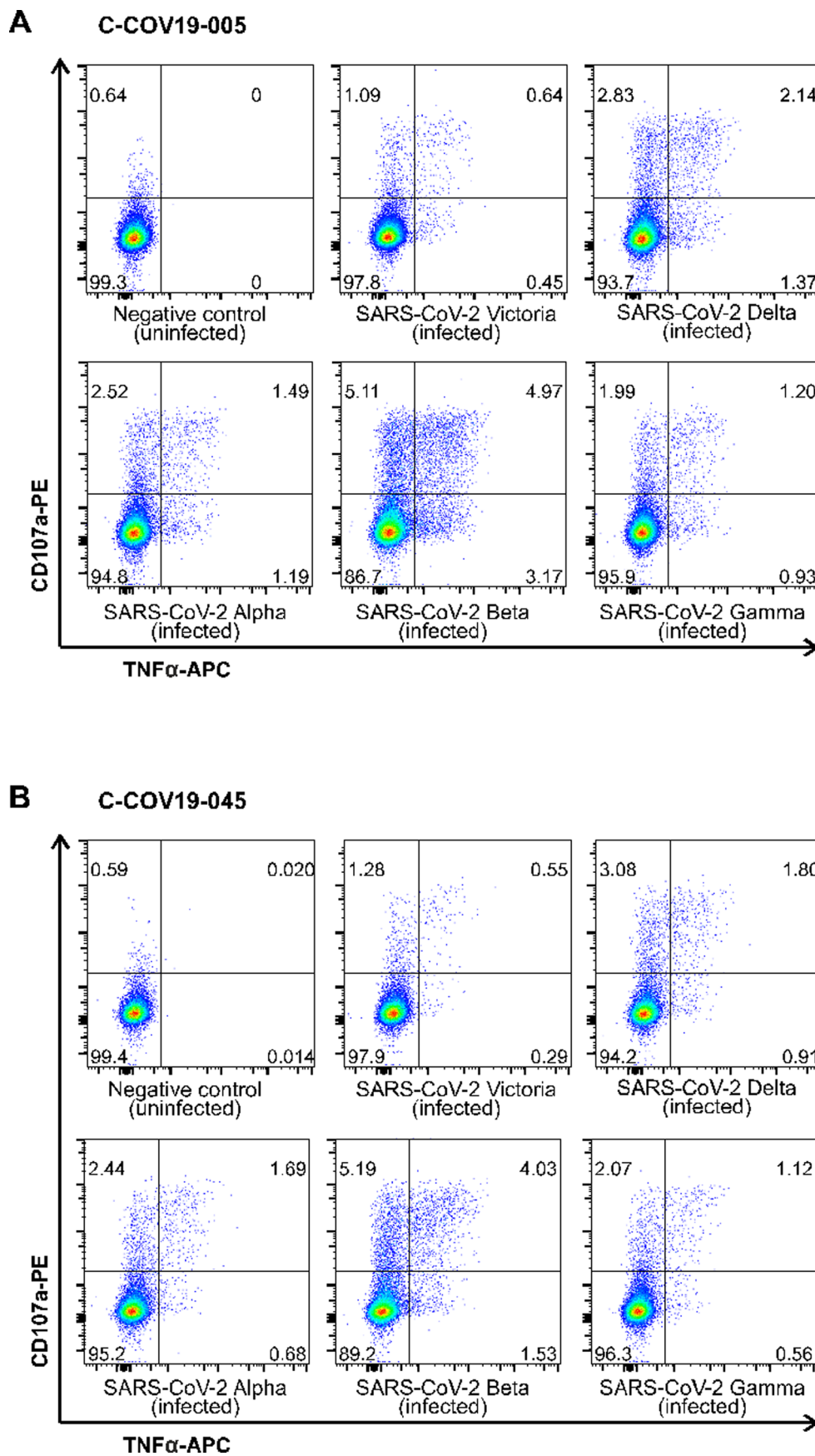
Extended Data Fig. 2 | TCR clonotypes for single cells and T cell clones and EC50 derivation. (a) Flow diagrams to show T cell clonotypes (defined as TRBV gene usage and CDR3 β sequence) between SmartSeq2 sequenced *ex vivo* single cells (*ex vivo* single T cells column) and bulk TCR sequenced T cell clones grown *in vitro* culture (T cell clones column). Panel on the left shows grouped TCRs from patients with mild disease, panel on the right for patients with severe disease. (b) Upper panel: IFN- γ ELISPOT assay for representative high and low functional avidity clones for each patient in blue and red respectively (C-COV19-038 only has low functional avidity clone shown). Lower panels: high functional avidity clone from C-COV19-45 and low functional avidity clone from C-COV-038 with example of EC50 derivation.



Extended Data Fig. 3 | FACS gating strategy and peptide titration. (a) Gating strategy for flow cytometric analyses. (b) Nucleoprotein peptide titration at varying concentrations for ICS for MIP1 β and CD107a. Representative flow cytometry plots shown (from single T cell clone).



Extended Data Fig. 4 | T cell response six months post infection to NP Vaccinia virus. Representative ICS flow cytometry plots showing MIP1β and CD107a protein expression after incubation with NP-expressing Vaccinia virus in three patient-derived bulk T cell lines from six months convalescent PBMC samples.



Extended Data Fig. 5 | T cell response six months post infection to SARS-CoV-2 and VOCs. Representative ICS flow cytometry plots showing TNF α and CD107a protein expression in patient-derived bulk T cell lines from six months convalescent samples after incubation with SARS-CoV-2 Victoria strain and VOCs (Alpha, Beta, Gamma and Delta) infected autologous B cell lines (A) patient C-COV19-005 and (B) C-COV19-045.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection ELISpot data were collected with AID ELISpot 7.0 and flow cytometry data were collected by BD FACSDiva V9.0/ Attune™ NxT software V3.2.1 Sequencing data were collected with NextSeq Control Software (NCS) version 4 or MiSeq Control Software Version 2.6.2.1

Data analysis Flow cytometry data were analyzed with FlowJo™ v.10.5.3 software for Mac. Prism version 8.2.1 for Mac OS. Raw BCL files from SmartSeq2 single cell RNA sequencing were converted to FASTQ format using bcl2fastq (v 2.20.0.422). FASTQ files were aligned to human genome hg19 using STAR (v 2.6.1D). Reads were counted using featureCounts (part of subread v 2.0.0). Resulting count matrices were analysed in R v 4.0.1 using Seurat (v 3.9.9.9010). Raw BCL files from 10x sequencing were processed using 10x Genomics Cellranger (v 5.0.0). Genetic demultiplexing was conducted using cellSNP (v 0.3.2) and Vireo (v 0.5.6). Packages used in R for additional analyses or data visualisation are: Harmony (v 1.0), ggplot2 (v 3.3.2), tcr (v 2.3.2), circlize (v 0.4.12), stringdist (v 0.9.6), ggalluvial (v 0.12.2), DESeq2 (v 1.28.1), pheatmap (v 1.0.12). TCR sequences were reconstructed from FASTQ files using MiXCR (v 3.0.13). GLIPH2 irtools.centos (v 0.01) run on a CentOS Linux platform (release 8.2.2011) was used for TCR analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The raw data from all the main figures and supplementary figures are available upon request. In addition the following published datasets were used: Lineburg et al (<https://doi.org/10.1016/j.immuni.2021.04.006>), Nguyen et al (<https://doi.org/10.1016/j.immuni.2021.04.009>) and COMBAT (<https://doi.org/10.1101/2021.05.11.21256877>)

Single cell raw data (excluding TCR sequences) can only be shared on request due to ongoing work for future publications.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	52 subjects recovered from COVID-19 samples (30 mild and 22 severe). This was a follow-up for a published study (Peng et al, NI, https://doi.org/10.1038/s41590-020-0782-6). Samples sizes were based on maximal available samples sets where detailed clinical and serological data were also available, and aligned well with our previously published data in other settings, including influenza and HIV (Lee et al, JCI 2008; Zhang et al, NC 2013; Zhao et al, AJCCM 2012)
Data exclusions	For ELISPOT assays, If negative control wells had >30 SFU/106 PBMCs or positive control wells (PHA stimulation) were negative, the results were excluded from further analysis. There were no data excluded from ELISpot analysis. For scRNA-Seq analysis cells were filtered using the following criteria: minimum number of cells expressing specific gene = 3, minimum number of genes expressed by cell = 200 and maximum number of genes expressed by cell = 4000. Cells were also excluded if they expressed more than 5% mitochondrial genes. For paired $\alpha\beta$ chain TCR analysis, cells were filtered to retain only $1\alpha1\beta$ or $2\alpha1\beta$ cells. For single β chain TCR analysis, cells were filtered to retain only 1β (regardless of number of α) to use for downstream analysis.
Replication	Samples analyzed in this study were from participants of a cohort study and samples were analyzed on individual study participants. Ex vivo experiments did not include replicates, but results from each participant were confirmed by at least two different experiments. For in vitro assays, results shown are always two or more independent experiments in which every repeat gave similar results.
Randomization	Randomization was not appropriate for this study of immune responses in COVID-19 convalescent individuals, with no associated therapeutic intervention
Blinding	Blinding was not appropriate for this study of immune responses in COVID-19 convalescent individuals, with no associated therapeutic intervention

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

Antibodies used for flow cytometry

Marker	Fluorophore	Supplier	Cat number	Clonotype	Lot Number	Dilution
CD14	BV510	BioLegend	367124	63D3	B280762	1:50
CD16	BV510	BioLegend	302048	3G8	B289732	1:50
CD19	BV510	BioLegend	302242	HIB19	B281769	1:50
CD8	PerCP-Cy5.5	BD Bioscience	565310	SK1	9312532	1:33
CD8	BV421	BioLegend	344748	SK1	B331004	1:40
CD3	FITC	BD Bioscience	345763	SK7	0342661	1:25
IFN γ	AF488	BD Bioscience	557718	B27	9128859	1:33
TNF α	APC	eBioscience	17-7349-82	MAB11	2101117	1:500
MIP1 β	APC-H7	BD Bioscience	561280	D21-1351	1046672	1:33
CD107a	PE	BD Bioscience	555801	H4A3	0073840	1:20
ACE2	Primary	R&D	AF933	Polyclonal	HOK0620051	1:20
anti-goat	AF647	AbCam	ab150135		GR3324428-3	1:1000

Validation

All antibodies used in this study are commercially available. Antibodies used in a specific species or application have been appropriately validated by manufacturers for that application and this information is provided on their website and product information datasheets. All antibodies described here have been further optimized for an appropriate concentration by testing several dilutions.

CD14 <https://www.biolegend.com/en-us/products/brilliant-violet-510-anti-human-cd14-antibody-14983>

CD16 <https://www.biolegend.com/en-us/products/brilliant-violet-510-anti-human-cd16-antibody-8003>

CD19 <https://www.biolegend.com/en-us/products/brilliant-violet-510-anti-human-cd19-antibody-8004>

TNF α <https://www.thermofisher.com/antibody/product/TNF-alpha-Antibody-clone-MAB11-Monoclonal/17-7349-82>

CD107a <https://wwwbdbiosciences.com/eu/applications/research/intracellular-flow/intracellular-antibodies-and-isotype-controls/anti-human-antibodies/pe-mouse-anti-human-cd107a-h4a3/p/555801>

CD3 <https://wwwbdbiosciences.com/sg/reagents/clinical/reagents/single-antibodies/cd3-fitc-sk7-also-known-as-leu-4/p/345763>

CD8 <https://wwwbdbiosciences.com/sg/reagents/research/antibodies-buffers/immunology-reagents/anti-non-human-primate-antibodies/cell-surface-antigens/percp-cy55-mouse-anti-human-cd8-sk1/p/565310>

MIP1 β <https://wwwbdbiosciences.com/sg/applications/research/intracellular-flow/intracellular-antibodies-and-isotype-controls/anti-human-antibodies/apc-h7-mouse-anti-human-mip-1-d21-1351/p/561280>

IFN γ <https://wwwbdbiosciences.com/sg/applications/research/t-cell-immunology/th-1-cells/intracellular-markers/cytokines-and-chemokines/human/alexa-fluor-488-mouse-anti-human-ifn-b27/p/557718>

CD8 <https://www.biolegend.com/en-gb/products/brilliant-violet-421-anti-human-cd8-antibody-13512>

ACE2 https://www.rndsystems.com/products/human-mouse-rat-hamster-ace-2-antibody_af933

donkey anti-goat AF647 <https://www.abcam.com/donkey-goat-igg-hl-alexa-fluor-647-preadsorbed-ab150135.html>

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

All the EBV-transformed B cell lines were established in the lab. Vero E6 cells (ATCC CCL-81), HEK293T cells (ATCC, CRL-11268), TK143B cells (ATCC, CRL-8303), rabbit RK13 cells (ATCC, CCL37) and African green monkey BS-C-1 cells (ATCC, CCL26) were from ATCC. Vero E6/TMPRSS2 cells were originally from NIBSC (reference 100978).

Authentication

Cell lines were validated by morphology. Vero E6 cells, HEK293T cells, TK143B cells, rabbit RK13 cells and BS-C-1 cells were obtained from ATCC and used at low passage. They were not further authenticated. Vero E6/TMPRSS2 from NIBSC were used for propagating SARS-CoV-2 virus and were not further authenticated. EBV transformed B cell lines expressing ACE2 were verified by flow cytometry staining of ACE2 and CD19.

Mycoplasma contamination

All the cell lines were tested negative for mycoplasma

Commonly misidentified lines (See [ICLAC](#) register)

No misidentified cell lines were used according to the version 11 of register of misidentified cell lines

Human research participants

Policy information about studies involving human research participants

Population characteristics

52 individuals were recruited following recovery from COVID-19, including 30 mild cases and 22 severe cases. Supplementary Figure 1 shows the participant characteristics. No significant differences in gender or age were noted between mild and severe groups, which was shown in previous publication (<https://doi.org/10.1038/s41590-020-0782-6>). The SaO₂/FiO₂ ratio in severe cases ranged from 4.3 (where 4.5 would be the estimate for an individual with mild disease breathing ambient air) to 1.6 with the patients with critical disease having an estimate of 0.8 (median in severe group 3.8).

Recruitment

Patients were recruited from the John Radcliffe Hospital in Oxford, UK, between March 2020 and April 2021 by identification of patients hospitalised during the SARS-CoV-2 pandemic. Patients were recruited into the Sepsis Immunomics study and had samples collected during their convalescence as well as during acute disease. For convalescence, patients were sampled at least 28 days from the start of their symptoms. Written informed consent was obtained from all patients. The potential bias, such as the timing when samples were taken, the gender and age of the patients, are unlikely to impact the results, as there is no significant difference in the age between two study groups, and no correlation was observed between the T cell

response and days post symptoms when samples were taken, as shown in previous publication (<https://doi.org/10.1038/s41590-020-0782-6>)

Ethics oversight

The samples were collected from patients with confirmed COVID who had consented to participate in Sepsis Immunomic project IRAS 260007. Ethical approval was given by the South Central-Oxford C Research Ethics Committee in England (Ref 19/SC/0296).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Cryopreserved PBMCs were thawed and rested overnight in R10 at 37°C. On the second day, PBMCs were stained with PE-conjugated HLA-B7 NP105-113 Pentamer (ProImmune, Oxford, UK). Live/Dead fixable Aqua dye (Invitrogen) was used to exclude non-viable cells from the analysis. Subsequently, cells were washed and stained with the following surface antibodies: CD3-FITC (BD Biosciences), CD8-PerCP-Cy5.5, CD14-BV510, CD19-BV510 and CD16-BV510 (Biolegend). After the final wash, cells were resuspended in 500 µl of PBS, 2 mM EDTA and 0.5% BSA (Sigma-Aldrich) solution and kept in dark at 4°C until flow cytometric acquisition. After exclusion of non-viable/CD19+/CD14+/CD16+ cells, CD3+ CD8+ Pentamer+ cells were sorted directly into 96-well PCR plates containing cell lysis buffer using a BD Fusion 1 or BD FACS Aria III (BD Biosciences). For intracellular cytokine staining (ICS), in vitro expanded T cell lines/clones were co-cultured with autologous B cell lines loaded with peptides, or infected with Vaccinia virus/live virus at an appropriate E:T ratio, in the presence of GolgiPlug (brefeldin A, BD), GolgiStop (monensin, BD) and 5 µl of PE-anti-CD107a (BD Biosciences) for 5 hrs. Then a standard FACS staining was carried out. Briefly, dead cells were first labelled with LIVE/DEAD™ Fixable Aqua dye and then followed by surface antibody staining. Subsequently, Cytotfix/Cytoperm™ kit (BD Biosciences) was used for permeabilizing the cells before staining the cells with antibodies against molecules expressed intracellularly. Finally, cells then be fixed with 1X cell fixing buffer.

Instrument

Samples were sorted using BD Fusion 1 or BD FACS Aria III (BD Biosciences) or acquired at Thermo Fisher Attune™ NxT Flow Cytometer

Software

Data were collected using FACS DIVA (v9.0.1) or Attune™ NxT software v3.2.1 and analyzed using FlowJo™ v.10 software for Mac.

Cell population abundance

All sorted samples were checked for after-sorting purity (>99%)

Gating strategy

For all the experiments, cells were first gated on single Lymphocytes by a forward side scatter gate. On sorting for RNASeq, after excluding dead cells, CD14+, CD19+, and CD16+ cells, HLA-B7-NP 105-113 specific T cells were identified as CD3+CD8+Pentamer+. For intracellular cytokine staining (ICS), the cytokine positive/ negative population were gated according to corresponding negative controls, known as unstimulated samples (T cells co-cultured with target cells without peptide loading or without virus infection): after excluding dead cells, cells then were gated into CD8+ T cells IFN-g+/-, TNFa+/-, MIP11β+/- and CD107a+/- populations were gated in consistence with the corresponding negative controls.

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.