Immunodominant NP105-113-B*0702 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease

Tao Dong (tao.dong@imm.ox.ac.uk)
University of Oxford https://orcid.org/0000-0003-3545-3758

Graham Ogg
University of Oxford

Yan-Chun Peng

Suet Felce
University of Oxford

Danning Dong
University of Oxford

Frank Penkava
University of Oxford

Alexander Mentzer
University of Oxford https://orcid.org/0000-0002-4502-2209

Xuan Lao
University of Oxford

Guihai Liu
University of Oxford

Zixi Yin
University of Oxford

Ji-Li Chen
University of Oxford

YongXu Lu
University of Cambridge

Dannielle Wellington
University of Oxford

Peter Wing
University of Oxford https://orcid.org/0000-0002-2354-3281

Delaney Dominey-Foy
University of Oxford

Julian Knight
University of Oxford

Chen Jin
University of Oxford
Wenbo Wang
University of Oxford
Megat Abd Hamid
University of Oxford
Ricardo Fernandes
University of Oxford
Beibei Wang
University of Oxford
Anastasia Fries
University of Oxford
Xiaodong Zhuang
Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ https://orcid.org/0000-0002-6870-9003
Neil Ashley
University of Oxford
Timothy Rostron
University of Oxford
Craig Waugh
University of Oxford
Paul Sopp
University of Oxford
Philip Hublitz
University of Oxford https://orcid.org/0000-0001-8810-3247
Ryan Beveridge
University of Oxford
Tiong Kit Tan
University of Oxford https://orcid.org/0000-0001-8746-8308
Christina Dold
University of Oxford
Andrew Kwok
University of Oxford
Charlotte Rich-Griffin
University of Oxford https://orcid.org/0000-0001-8212-9542
Wanwisa Dejnirattisai
University of Oxford
Isar Nassiri
University of Oxford

Robert Watson
University of Oxford
Orion Tong
University of Oxford https://orcid.org/0000-0002-0659-5944
Chelsea Taylor
King's College London
Piyush Kumar Sharma
University of Oxford https://orcid.org/0000-0002-6247-3503
Fabiola Curion
University of Oxford
Santiago Revale
University of Oxford
Lucy Garner
University of Oxford https://orcid.org/0000-0002-6461-252X
Kathrin Jansen
University of Oxford https://orcid.org/0000-0002-3803-7680
Ricardo Ferreira
University of Oxford
Moustafa Attar
University of Oxford
Jeremy Fry
ProImmune Limited
Rebecca Russell
University of Oxford
Hans Stauss
University College London https://orcid.org/0000-0003-4340-7911
William James
william.james@path.ox.ac.uk https://orcid.org/0000-0002-2506-1198
Alain Townsend
University of Oxford
Ling-Pei Ho
University of Oxford
Juthathip Mongkolsapaya
University of Oxford https://orcid.org/0000-0003-3416-9480
Paul Klenerman
University of Oxford https://orcid.org/0000-0003-4307-9161
Gavin Screaton
University of Oxford https://orcid.org/0000-0002-3549-4309
Calliope Dendrou

University of Oxford
Stephen Sansom
University of Oxford https://orcid.org/0000-0003-4569-8513
Rachael Bashford-Rogers
University of Oxford https://orcid.org/0000-0002-6838-0711
Benny Chain
UCL https://orcid.org/0000-0002-7417-3970
Geoffrey Smith
University of Cambridge https://orcid.org/0000-0002-3730-9955
Jane McKeating
University of Oxford
Benjamin Fairfax
University of Oxford
Paul Bowness
Oxford University Hospitals https://orcid.org/0000-0003-4597-0484
Andrew McMichael
University of Oxford

Article

Keywords: SARS-CoV-2, COVID-19, immunology, vaccine development

Posted Date: July 27th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-734011/v1

Version of Record: A version of this preprint was published at Nature Immunology on December 1st, 2021. See the published version at https://doi.org/10.1038/s41590-021-01084-z.

Abstract

$NP_{105-113}$-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 $NP_{105-113}$-B*07:02 specific T-cell clones and single cell sequencing were performed concurrently, with functional avidity and anti-viral efficacy assessed using an *in vitro* SARS-CoV-2 infection system, and were correlated with TCR usage, transcriptome signature, and disease severity (acute N=77, convalescent N=52). We demonstrated a beneficial association of $NP_{105-113}$-B*07:02 specific T-cells in COVID-19 disease progression, linked with expansion of T-cell precursors, high functional avidity and anti-viral effector function. Broad immune memory pools were narrowed post-infection but $NP_{105-113}$-B*07:02 specific T-cells were maintained 6 months after infection with preserved anti-viral efficacy to the SARS-CoV-2 Victoria strain, as well as new Alpha, Beta and Gamma variants. Our data shows that $NP_{105-113}$-B*07:02 specific T-cell responses associate with mild disease and high anti-viral efficacy, pointing to inclusion for future vaccine design.

Introduction

$CD8^+$ 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 HLA types [3, 4, 5, 6]; in particular, we found multiple dominant epitopes in NP (nucleoprotein) 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 M (membrane) $CD8^+$ T cell responses are associated with mild disease and NP is one of the most common targets for $CD8^+$ dominant T cell responses in SARS-CoV-2 infection [3].

Among the dominant epitopes identified to date, $NP_{105-113}$-B*07:02 appears to be among the most dominant [3, 4, 6, 7]; notably, no variants are found within this epitope from over 300k sequences in COG-UK global sequence data alignment [8]. 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 neutralising antibodies, in order to protect against newly emergent viral strains that escape antibody responses to spike in some cases. [9].

Recently, Lineburg and colleagues showed a biased TRBV27 gene usage, with long CDR3β loops preferentially expressed in $NP_{105-113}$-B*07:02 specific TCRs in both unexposed and COVID-19 recovered individuals [10]. This 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 by Nguyen and colleagues 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*0702-expressing individuals [7].

In this study, we present an in-depth analysis to explore correlations between $NP_{105-113}$-B*07:02 specific T cell responses, T cell receptor (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*0702 $NP_{105-113}$ 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 to severe cases. Single cell analysis revealed that preserved beneficial functional phenotypes are associated with protection from severe illness and have better overall anti-viral function. In addition, $NP_{105-113}$-B*07:02 specific T cells can recognise the naturally processed epitope in live virus and recombinant Vaccinia virus infected cells, which correlates with anti-viral efficacy.

Results

$NP_{105-113}$-B*07:02 specific T cell responses are stronger in patients who have recovered from mild COVID-19 infection

In our previous study, we identified five dominant $CD8^+$ epitopes targeting NP, including the most dominant epitope $NP_{105-113}$ (amino acid sequence SPRWYFYYL) restricted by HLA-B*07:02 [3]. 52 individuals who recovered from COVID-19 were included in this current study, comprising 30 mild cases and 22 severe cases (including 4 with critical illness, STAR Methods). All the patients were HLA typed and 19 (36.5%) were HLA-B*07:02 positive (10 mild and 9 severe cases). We proceeded to carry out e*x vivo* IFN-γ ELISpot assays using HLA-B*0702 positive convalescent samples 1-3 months post infection (STAR Methods). 79% (15/19) of HLA-B*07:02 individuals showed responses to this epitope (Figure 1A), further confirming the dominance of this T cell response in our cohort. Among these responders, we found a higher proportion (60%) had recovered from mild infection (Figure 1B) and made significantly stronger responses to this epitope, compared to those who had recovered from severe disease (Figure 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% (Figure 1D, P=0.015). In addition, we did not find HLA-B*0702 association with disease outcome in our study cohorts (Figure 1E, 77 acute and 52 convalescent patients). Our data

highlights the association of this dominant epitope-induced T cell response with mild disease outcome and provides 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 MHC-class I pentamers using flow cytometry. We performed single cell analysis using SmartSeq2 (STAR Methods) for PBMC samples from four convalescent patients, including two who recovered from mild COVID-19 infection (C-COV19-005 and C-COV19-046) and two who recovered from severe disease in early infection (C-COV19-038 and C-COV19-045). TCR sequences and transcriptomic profiles of each single cell were analysed.

We compared gene expression of 348 $CD8^+$ NP-specific sorted single cells isolated from mild (N=208 from two patients) and severe cases (N=140 from two patients) by scoring expression levels of manually defined gene sets (Supplementary Table 1). Gene signatures associated with T cell cytotoxicity and inhibitory receptors were analysed and compared between severity groups (STAR Methods). We found that cells from patients who had recovered from severe COVID-19 have significantly higher cytotoxicity gene expression scores (Figure 2A, P=0.00032), with upregulation of *GZMK* (P=3.02E^{-05}) and *GNLY* (P=1.41E^{-09}) (encoding granzyme K and granulysin respectively, Figure 2B). These cells also displayed increased inhibitory receptor expression (Figure 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 are 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], we found that $NP_{105-113}$-B*07:02-specific T cells from our cohort show very broad TCR repertoires. The circos plots in Figure 3A show paired TCR α and β chains (V and J gene usage) from the four individuals analysed with SmartSeq2 single cell RNA-Seq. Figure 3B shows the combined TCR repertoire of all four patients represented by TCR clonotype (defined separately for each patient combining V gene and CDR3 amino acid sequence). Although the $NP_{105-113}$-specific TCR repertoire is diverse with unique pairings of Vα and Vβ genes, we observed that 15/45 (33.3%) of unique Vβ clonotypes were paired with several distinct Vα clonotypes. In contrast, there is only 1/55 (1.8%) Vα clonotype that pairs to multiple Vβ clonotypes; this highlights the importance of studying Vβ in the TCR repertoire. Further detailed TCR information can be found in Supplementary Table 2.

CDR3β sequences of HLA-B*07:02 $NP_{105-113}$-specific T cells from patients with mild COVID-19 display higher similarity to naïve precursors found in pre-pandemic individuals

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]. Both studies found TCRs specific to $NP_{105-113}$-B*07:02 in SARS-CoV-2 unexposed and infected individuals. Nguyen *et al*. revealed that these cells are likely to be naïve. This is very different from the central/effector memory phenotype of SARS-CoV-2 specific T cells reported by ourselves and others. Overall, our data supported the notion that T cells bearing TCRs specific to $NP_{105-113}$-HLA-B*07:02 in SARS-CoV-2 unexposed individuals are thought unlikely to have resulted from previous seasonal coronavirus infection [7]. 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.

To take advantage of the results from our SmartSeq2 single cell RNA-Seq, we first compared TCR sequences from our four convalescent COVID-19 patients to pre-pandemic TCR sequences from healthy donors published by Lineburg, Nguyen and another study cohort, COMBAT [14]. 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 $NP_{105-113}$ epitope, however for the COMBAT dataset, we used GLIPH2 analysis [15] to extract TCRs with predicted specificity to this epitope based on convergence with known $NP_{105-113}$-specific TCRs (STAR Methods).

We calculated similarity scores for CDR3β amino acid sequences between pairwise combinations of SmartSeq2 TCRs and pre-pandemic/healthy TCRs (STAR Methods). A similarity score of 1 indicates that the pair of CDR3β sequences are identical, while a score of 0 indicates complete dissimilarity. In our convalescent patient cohort, CDR3β from patients with mild disease are more similar to TCRs from pre-pandemic/healthy individuals, than those from severe patients (Figure 4A, $P<2.20E^{-16}$). To further support our findings, we looked at the proportion of TCR sequences from mild and severe patients (both acute and convalescent) that can be found in the same convergence groups as sequences from healthy donors, indicating high CDR3β 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 (Figure 4B, $P<2.2E^{-16}$).

We were able to link predicted $NP_{105-113}$-B*07:02 TCRs 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 $NP_{105-113}$ specificity. Cellular subtyping of these $CD8^+$ $NP_{105-113}$-B*07:02 T cells show a higher proportion of naïve T cells in one HLA-B*07:02 healthy individual compared to predominantly T effector memory subtypes in acute COVID-19 patients (N=17, Figure 4C). This reinforces the finding that only $NP_{105-113}$-B*07:02 specific T cells from acute HLA-B*07:02 positive patients are exposed to antigen and undergo T

cell differentiation, whereas $NP_{105-113}$-specific T cells in pre-pandemic individuals are naïve precursors rather than memory cells from prior cross-reactive infection.

$NP_{105-113}$-B*07:02-specific T cells have a broad range of functional avidity and high functional avidity is associated with clonotype expansion in mild disease

In parallel with single cell sorting for SmartSeq2, we also sorted, cloned and expanded $NP_{105-113}$-B*07:02-specific T cells from the same convalescent COVID-19 patients *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 3. 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 (Figure S1A).

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 TRBVβ gene usage and CDR3β sequence (Table 1). T cell functional avidity was measured by IFN-γ ELISPOT and calculated by EC50 (Figure S1B, Supplementary Table 4). We found evidence for low and high functional avidity groups (Figure 5A) based on the EC50 of T cell clones, with EC50 < 0.11 considered as high avidity, and those with EC50 > 0.11 are low avidity T cells (STAR Methods). We then aggregated RNA counts from single cells (pseudobulk) to compare differences in gene expression between the two avidity groups. Although there were only 7 significantly differentially expressed genes (Figure 5B), possibly as a result of small samples sizes and patient variation, differentially expressed genes of note upregulated in high functional avidity cells include *IL10RA*, *PARK7* and *LTA4H*. The interaction of IL-10 with IL10RA expressed on $CD8^+$ T cells has been reported to directly decrease $CD8^+$ T cell antigen sensitivity in patients with chronic hepatitis C (HCV) infection [18], while *PARK7* is promotes survival and maintains cellular homeostasis in the setting of intracellular stress [19]. LTA4H is an enzyme with known potent anti-inflammatory activity, and 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 illness [20]. 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 (Figure 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 $NP_{105-113}$-B*07:02-specific T cells responding to naturally processed epitope is correlated 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 Epstein-Barr virus (EBV) transformed B cells with Vaccinia virus expressing NP and co-cultured with $NP_{105-113}$-B*07:02-specific T cell clones. T cell degranulation and cytokine production was then assessed by intracellular staining after six hours of incubation (STAR Methods). Figure 6A shows an example of CD107a expression and MIP1β chemokine production from a representative T cell clone. Gating for CD107a and/or MIP1β producing cells were based on corresponding negative controls (Figure S2A). When compared to the peptide-loaded targets, we found that the response to Vaccinia virus infected BCLs was much weaker, consistent with lower antigen loads. As shown in Figure S2B, the load of this naturally processed and presented epitope was equivalent to no more than 3nM peptide. Nevertheless, NP Vaccinia virus-incubated clones with high CD107a expression showed a negative correlation with their individual EC50 values (Figure 6B, 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 MIP1β producing cells (Figure 6C, R=-0.6879, *P*=0.0082).

To further investigate the anti-viral activity of $NP_{105-113}$-B*07:02 specific T cells, we established an *in vitro* SARS-CoV-2 infection system (STAR Methods). Briefly, the *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 FACS sorting and maintained by antibiotic selection, after which cells were subsequently used for SARS-CoV-2 virus infection. After 48 hours incubation, intracellular viral copies were quantified by quantitative PCR, where the reduction of virus replication is calculated as percentage of virus suppression by T cells. Figure 6D shows the suppression of virus replication by each individual T cell clone (SARS-CoV-2 Victoria strain). We found that the percentage of virus suppression was strongly correlated with their functional avidity; therefore, high avidity T cells can efficiently inhibit viral replication (Figure 6E, R=-0.7699, *P*=0.0075).

$NP_{105-113}$-B*07:02 specific T cells are maintained six months after infection with proportionally narrowed TCR repertoire and preserved anti-viral efficacy

In order to examine whether the memory T cells established post-natural infection could provide sufficient protection against secondary viral infection, we collected PBMCs from three patients (C-COV19-005, C-COV19-045, C-COV19-046) six months after infection and sequenced sorted $CD8^+$ $NP_{105-113}$-B*07:02-specific T cells. We discovered that six 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 (Figure 7A). We then isolated and expanded further $NP_{105-113}$-B*07:02-specific T cell bulk lines from PBMC samples taken six months after infection. We assessed the anti-viral efficacy of these bulk T cell lines in our *in vitro* SARS-CoV-2 infection assays: all three T cell lines showed increased MIP1β and CD107a protein expression after incubation with NP-expressing Vaccinia virus

(Figure S3) and SARS-CoV-2 infected BCLs (Figure 7B), and are capable of suppressing SARS-CoV-2 replication (Figure 7C). We further found that these antigen-specific bulk cell lines showed strong inhibition against current variants of concern, including the recently emerged Alpha, Beta and Gamma SARS-CoV-2 variants (Figure 7D). 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, 4, 5, 6]. We found high similarity of TCRs in both HLA-B*07:02 positive and HLA-B*07:02 negative COVID-19 recovered individuals, with naive precursors identified in pre-pandemic 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 anti-viral function compared with convalescent severe COVID-19. Interestingly, the immune memory pools post-infection (six months convalescence) are narrowed but remain proportional; we found no bias towards 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*0702 when analysed alone.

The highly diverse T cell receptor 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 if 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 pre-pandemic individuals, in particular patients with mild symptoms. We hypothesise that $NP_{105-113}$-B*07:02 specific T cell responses play an important role in protecting individuals from severe illness, which is likely due to early priming and expansion of high frequency naive TCRs specific to this epitope.

We further provided evidence to support our hypothesis by studying a cohort of patients with acute SARS-CoV-2 infection, by analysing the TCR repertoire in HLA-B*07:02 positive patients. We first found high frequencies of TCR precursors with naïve phenotype in HLA-B*07:02 positive healthy donors; this further supports the recent findings from Nguyen *et al.*, 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 is associated with disease severity, where $NP_{105-113}$-B*07:02 specific T cells are more activated and well differentiated in individuals recovered from severe illness. This is likely as 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 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 [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 $NP_{105-113}$-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 response to naturally processed epitopes presented by infected cells is correlated 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 six months post-infection), 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, firstly, high frequency naive T cell precursors recognising $NP_{105-113}$-B*07:02 that are observed in pre-pandemic individuals are HLA-B*07:02 independent; secondly, the protective effect of $NP_{105-113}$-HLA-B*07:02 specific TCRs from severe illness may be due to early expansion of high frequency naïve T cell precursors bearing these TCRs. Moreover, we found that the TCR repertoire is not disturbed following 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.

Limitations of the study

We recognise that the number of convalescent patients analysed by single cell gene expression and TCR sequencing (N=4) is small. Also, the number of $NP_{105-113}$-B*07:02-specific cells from pre-pandemic donors and acute COVID-19 patients is low; this is partly because these cells were not pentamer-sorted before analysis. In this study we focus on $CD8^+$ T cell responses to a single epitope, however it may be useful in the future to see if there are any shared or distinct features with other dominant responses. The specifics of antigen loading of this particular epitope compared with other NP epitopes, as well as variation in levels of protein expression and localisation is also unknown, and warrants further investigation.

Materials And Methods

Study participants

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. Patients were sampled at least 28 days from the start of their symptoms. Written informed consent was obtained from all patients. Ethical approval was given by the South Central-Oxford C Research Ethics Committee in England (Ref 13/SC/0149).

Clinical definitions

All patients were confirmed to have a positive test for SARS-CoV-2 using reverse transcriptase polymerase chain reaction (RT-PCR) from an upper respiratory tract (nose/throat) swab tested in accredited laboratories. The degree of severity was identified as mild, severe or critical infection according to recommendations from the World Health Organisation. Patients were classified as 'mild' if they did not require oxygen (that is, their oxygen saturations were greater than 93% on ambient air) or if their symptoms were managed at home. A large proportion of our mild cases were admitted to hospital for public health reasons during the early phase of the pandemic even though they had no medical reason to be admitted to hospital. Severe infection was defined as COVID-19 confirmed patients with one of the following conditions: respiratory distress with RR>30/min; blood oxygen saturation<93%; arterial oxygen partial pressure (PaO2) / fraction of inspired O2 (FiO2) <300mmHg; and critical infection was defined as respiratory failure requiring mechanical ventilation or shock; or other organ failures requiring admission to ICU. Since the Severe classification could potentially include individuals spanning a wide spectrum of disease severity ranging from patients receiving oxygen through a nasal cannula through to non-invasive ventilation, we also calculated the SaO2/FiO2 ratio at the height of patient illness as a quantitative marker of lung damage. This was calculated by dividing the oxygen saturation (as determined using a bedside pulse oximeter) by the fraction of inspired oxygen (21% for ambient air, 24% for nasal cannula, 28% for simple face masks and 28, 35, 40 or 60% for Venturi face masks or precise measurements for non-invasive or invasive ventilation settings). Patients not requiring oxygen with oxygen saturations (if measured) greater than 93% on ambient air, or managed at home were classified as mild disease. Viral swab Ct values were not available for all patients. In addition, we have standardised all of our analyses to the days since symptom onset.

Generating ACE2 transduced Epstein-Barr virus (EBV)-transformed B cell lines.

Epstein-Barr virus (EBV)-transformed B cell lines (BCLs) were generated as described previously [26]. The cDNA for the human Angiotensin Converting Enzyme 2 (*ACE2*) gene (ENSG00000130234) was cloned into a lentiviral vector that allows co-expression of eGFP and a Puromycin resistance marker (Addgene, Plasmid 17488). The plasmids were co-transfected with the packaging plasmids pMD2.G and psPAX2 into HEK293-TLA using PEIpro (Polyplus). Lentiviral supernatant was collected 48h and 72h post transfection and concentrated by ultracentrifugation. EBV-transformed BCLs were infected by ACE2-lentivirus at MOI 0.1 with the addition of 8µg/ml polybrene (Sigma-Aldrich, Cat. # TR-1003-G) overnight, then washed, and further cultured for 3 - 5 days. ACE2-expressing B cells were stained using a primary

goat anti-human ACE2 antibody (R&D, AF933) and a donkey anti-goat AF647 secondary antibody (Abcam, Cat. # ab150135) followed by cell sorting by flow cytometry. B cells with stable expression of ACE2 were maintained with 0.5μg/ml puromycin (ThermoFisher, Cat. # A1113803). Mycoplasma test was caried out every four weeks with all the cell lines using Lonza MycoAlert detection kit (Lonza,

Generating T cell lines and clones

Short-term SARS-CoV-2-specific T cell lines were established as previously described [17]. Briefly, 3×10^6 to 5×10^6 PBMCs were pulsed as a pellet for 1 h at 37°C with 10μM of peptides containing T cell epitope regions and cultured in R10 (RPMI 1640 medium with 10% foetal calf serum (FCS), 2mM glutamine and 100mg/ml pen/strep) at 2×10^6 cells per well in a 24-well Costar plate. IL-2 was added to a final concentration of 100U/mL on day 3 and cultured for a further 10 - 14 days. 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 previously [27].

Generating Vaccinia virus expressing SARS-CoV-2 NP

SARS-CoV-2 nucleocapsid (NP) expression vectors (gifts from Dr. Peihui Wang, Shandong University, Shandong, China [28]) were first digested with *KpnI* and SacII. The resulting fragment was then cloned into VACV expression vector pSC11, which was inserted with a DNA segment encoding *KpnI* and *Sac*II digestion sites (GGTACCGCGGCCGCCCGCGG). The SARS-CoV-2 NP-expressing recombinant Vaccinia virus (rVACV) was produced as described previously. [29, 30, 31]. In brief, HEK293T cells (ATCC, CRL-11268) were transfected with 3μg of pSC11 containing NP in the presence of polyethylenimine. At 24h post transfection, cells were infected with the Lister strain of VACV at MOI 1 for 48h. The infected cells were collected for recombinant virus purification using TK143B cells (ATCC, CRL-8303) in the presence of 25μg/ml bromodeoxyuridine. The NP-expressing rVACV was selected through β-galactosidase staining by supplementing 25μg/ml X-gal to an agarose overlay. Master stocks of rVACV were prepared by infection on rabbit RK13 (ATCC, CCL37) and titrated on African green monkey BS-C-1 (ATCC, CCL26) cells.

SARS-CoV-2 live virus propagation and titration.

SARS-CoV-2 Victoria 01/20 strain (BVIC01), and variants of concerns: Alpha (Lineage B1.1.7, 20I/501Y.V1.HMPP1) and Beta (Lineage B1.351, 201/501.V2.HV001) were originally from Public Health England and provided by Prof. Jane McKeating [32]. SARS-CoV-2 Gamma (Lineage P.1) was provided by Prof. Gavin Screaton [33]. Virus was propagated and titrated as previously described. [32, 33]. In brief, Victoria 01/20, Alpha and Beta were propagated with Vero E6 cells, whereas SARS-CoV-2 Gamma was propagated with Vero E6/TMPRSS2 (kindly provided by Alain Townsend). Naïve Vero E6 or Vero E6/TMPRSS2 cells were plated one night before and infected with SARS-CoV-2 at MOI of 0.003. The cultures were harvested when visible cytopathic effect was observed 48 - 72h after infection and residual cell debris was removed by centrifugation. The virus-containing supernatant was aliquoted and stored at

−80°C. Viral titre was determined by plaque assay with Vero E6 or Vero E6/TMPRSS2 as previously described, and plaque-forming units (PFU) per mL was used to calculate 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 previously [3]. For *ex vivo* ELISpots, peptides were added to 200,000 PBMCs per test at the final concentration of 2μg/mL for 16–18 h. When using T cell clones, autologous EBV-transformed B cell lines were first loaded with peptides at 3-fold titrated concentrations and were subsequently co-cultured with T cells at an effector: target (E:T) ratio of 1:50 for at least 6h. To quantify antigen-specific responses, mean spots of the control wells were subtracted from the positive wells (PHA stimulation), and the results expressed as spot forming units (SFU)/10^6 PBMCs. Responses were considered positive if results were at least three times the mean of the negative control wells and >25SFU/10^6 PBMCs. If negative control wells had >30SFU/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 PE-conjugated HLA-B7 $NP_{105-113}$ 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 500μl 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 (Thermo Fisher, UK) using a BD Fusion sorter or BD FACS Aria III (BD Biosciences) and stored at −80 °C for subsequent analysis.

Single cell RNA Sequencing (RNA-Seq)

Single cell RNA-Seq with ex vivo sorted $CD8^+Pentamer^+$ T cells was performed using SmartSeq2 [34]. with following modifications. Reverse-transcription and PCR amplification were performed as described [34] with the exception of using ISPCR primer with biotin-tagged at 5' 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.

Deep sequencing of T cell receptor (TCR) repertoire of T cell clones

100,000 cells from each T cell clone were harvested and washed three times with PBS. Total RNA was extracted using RNeasy Plus Micro Kit (Qiagen, Cat. # 74034). 100ng of total RNA from each of T cell clone was used to generate full length of TCR repertoire libraries for Illumina Sequencing using SMARTer Human TCR a/b Profiling Kit (Takara, Cat. # 635016) following supplier's instructions. In brief, first-strand

cDNA was synthesised by reverse transcription using TRBC reverse primers and further extended with a template-switched SMART-Seq®v4 Oligonucleotide. Following reverse transcription, two rounds of PCR were performed in succession to amplify cDNA sequences corresponding to variable regions of TCR-α and/or TCR-β transcripts with primers including Illumina indexes allowing for sample barcoding. PCR products were then purified using AMPure beads (Beckman Coulter). The quantity and quality of cDNA libraries were checked on Agilent 2100 Bioanalyzer system. Sequencing was performed with MiSeq reagent Kit v3 (600 cycles) on MiSeq (Illumina).

Intracellular cytokine staining (ICS)

Intracellular cytokine staining was performed as described previously [3]. Briefly, T cells were co-cultured with peptide-loaded or virus-infected BCLs at an appropriate E:T ratio for a 6h incubation with GolgiPlug and GolgiStop, and surface stained with PE-anti-CD107a. Dead cells were labelled using LIVE/DEAD™ Fixable Aqua dye (Invitrogen); after staining with BV421-anti-CD8, cells were then washed, fixed with Cytofix/Cytoperm™ and stained with AF488-anti-IFNγ, APC-anti-TNFα (eBioscience) and APC-H7-anti-MIP1ß. 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 and analyzed using FlowJo™ v.10 software (FlowJo LLC).

Evaluation of T cell response to Vaccinia virus infection

EBV transformed B cell lines were infected with Lister strain Vaccinia virus at a MOI 3 for 90-120 mins at 37°C. Cells were then washed three times to remove virus and incubated overnight in R10 at 37°C. The next day, cells were counted and co-cultured with T cells at an E:T ratio of 1:1. Degranulation (CD107a expression) and cytokine production of T cells were then evaluated by ICS as described above.

Evaluation of T cell response to live virus infection

EBV transformed BCLs expressing ACE2 were infected SARS-CoV-2 viruses at a MOI 3 for 120 mins at 37°C. Cells were then washed three times and incubated in R10 at 37°C. After 48h, cells were then counted and co-cultured with T cells at an E:T ratio of 1:1. Degranulation (CD107a expression) and cytokine production of T cells were then evaluated by ICS as described above.

Live Virus Suppression Assay (LVSA)

EBV transformed BCLs expressing ACE2 were infected with SARS-CoV-2 viruses at MOI 0.1 for 120 mins at 37°C. Cells were then washed three times and co-cultured with T cells at an E:T ratio of 4:1. Control wells containing virus-infected targets without T cells were also included. After 48hrs incubation, the cells were washed with PBS and lysed with Buffer RLT (Qiagen). RNA was extracted using RNeasy 96 kit (Qiagen, Cat. # 74181). Virus copies were then quantified with Takyon™ Dry one-step RT-qPCR (Eurogentec, Cat. # UFD-NPRT-C0101) using SARS-CoV-2 (2019-nCoV) CDC qPCR Probe Assay (IDT, ISO

13485:2016) and human B2M (Beta-2-Microglobulin) as an endogenous control (Applied Biosystems™). The suppression rate was calculated by the percentage of reduction of virus replication by T cells.

SmartSeq2 single cell data processing

BCL files were converted to FASTQ format using bcl2fastq version 2.20.0.422 (Illumina). FASTQ files were aligned to human genome hg19 using STAR version 2.6.1d [35]. Reads were counted using featureCounts (part of subread version 2.0.0, [36]). The resulting counts matrix was analysed in R version 4.0.1 using Seurat version 3.9.9.9010 [37].

SmartSeq2 single cell RNA sequencing analysis

After creating a Seurat object from the counts matrix, 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. Patient-specific cells were integrated using Harmony version 1.0 to remove batch effects (differences between patients) but to still retain meaningful biological variation. The *AddModuleScore* function from the Seurat package was used to look at expression of specific gene sets (Supplementary Table). 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 to the rest of the population. Module scores were plotted using ggplot2 version 3.3.2 [38].

SmartSeq2 TCR repertoire analysis

TCR sequences were reconstructed from RNA sequencing FASTQ files using MiXCR version 3.0.13 [39, 40] and the command *mixcr analyze shotgun* to produce separate TRA and TRB output files for analysis. The output text files were parsed into R using tcR version 2.3.2. For paired αβ TCRs, cells were filtered to retain only 1α1β or 2α1β cells. Circos plots showing paired αβ TCRs were created using circlize version 0.4.12 [41]. Separately, lists were generated for all 1β cells (regardless of number of α) to use for downstream analysis.

Clustering

The input data for clustering was all 1β from SmartSeq2 single cells and 1β from bulk sequencing T cell clones. Single cells and clones were grouped by Vβ usage first; TCRs from either single cells or clones which had unique Vβ gene usage were excluded. Each Vβ group was broken down into subgroups based on CDR3β sequence; any TCRs from either single cells or clones that contained unique CDR3β sequences were excluded. Only TRBV27, TRBV28, TRBV5-1 showed multiple CDR3β sequences with the same gene usage. After plotting the EC50 values of the 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 Vβ

gene usage and CDR3β sequences shared among the TCRs from single cell sequencing and bulk T cell clone sequencing.

In order to group as many of the SmartSeq2 single cells into one of these 18 groups, the *stringsim* function was used from the stringdist package version 0.9.6 [42] to compare the similarity between all SmartSeq2 CDR3β sequences and each of the 18 CDR3β from the single cell/clone grouping. A minimum similarity score of 0.7 was used to decide if a TCR from a SmartSeq2 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 converted to FASTQ files as before. TCRs reconstructed using MiXCR command *mixcr analyze amplicon*, and the resulting output files (TRA and TRB) were parsed into R using tcR as before. TCRs were filtered to retain only 1α1β for each clone. TCR clonotypes (defined as Vβ gene usage and CDR3β sequence) were compared between single TCR and bulk TCR sequencing using using ggalluvial version 0.12.2 [43]. The predicted functional avidity annotation was overlayed onto the plots using the *stringsim* function as previously described to classify TCRs into high or low functional avidity groups (minimum score 0.5).

10X VDJ sequencing

Raw files were processed using 10x Genomics Cellranger version 5.0.0 [44]. BCL files were converted to FASTQ format using *cellranger mkfastq* and TCRs were reconstructed using *cellranger vdj*. To carry out donor deconvolution from multiplexed single cell data, cellSNP version 0.3.2 [45] was first used to generate a list of SNPs from cellranger output (BAM file). Subsequently Vireo version 0.5.6 [46] was used to demultiplex the sequencing data into individual patients from the pooled sequenced libraries, based on the previously generated SNPs list. The filtered_contig_annotations.csv file was annotated with patient information and used for subsequent downstream analysis. TCRs from 10X sequencing represent six months convalescence and were compared to one month convalescence TCRs (SmartSeq2) from the same patient using ggalluvial. The predicted functional avidity annotation was overlayed onto 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

Normalised 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 (/CBD-10X-00010/multimodal.annotation.release.1.0.tsv); COMBAT TCR chain information (CBD-10X-00008/tcr_chain_information.tsv.gz); and patient metadata (/CBD-CLINNORM-00006/COMBAT_clinical_basic_data_freeze_160221.txt). 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 WHO 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 either Uninfected (0), Mild (1-4), Severe (5-7) or Death (8).

GLIPLH2 analysis

A GLIPH2 $CD8^+$ TCR input file was created from the following datasets: COMBAT 10x paired chain single cell and bulk TCR from all available participants [14]; 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 Lineburg and Nguyen datasets [7, 10]. Clonotypes were defined as having a unique combination of CDR3β amino acid sequence, TRBV gene, TRBJ gene and CDR3α amino acid sequence. Where no or multiple CDR3α sequences were available for a cell, an NA value was used for the CDR3α 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 10x 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 [3, 14], in addition to published HLA data relating to the Lineburg and Nguyen datasets [7, 10].

The GLIPH2 irtools.centos version 0.01 [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 [http://50.255.35.37:8080/] 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; 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 of <0.05 or >=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 HLA-B*07:02 positive TCRs. Any clonotypes from convergence groups lacking HLA-B*07:02 association but belonging to patients having 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 COMBAT dataset or experimentally defined from Lineburg *et al*. and Nguyen *et al*.) were compiled to form a single list of sequences (237 TCRs). The function *stringsim* was used to generate similarity scores from pairwise comparisons between each CDR3β sequence from the pre-pandemic/healthy list and each CDR3β sequence from 85 unique clonotypes from four convalescent COVID-19 patients (clonotype defined per patient, TRBV gene usage and CDR3β sequence). A score of 1 indicates total similarity while a score of 0 is total dissimilarity. Each score was plotted on a box plot using ggplot2 version 3.3.2.

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 object version 1.10.1 [47]. Differential gene expression was conducted using DESeq2 version 1.28.1 on aggregated (pseudobulk) counts. Significant genes were visualised on a heatmap using pheatmap version 1.0.12.

Statistics

Mann Whitney non-parametric test to compare two groups (R Studio); other statistical tests carried out using GraphPad Prism. Nonlinear regression with variable slope (four parameters) in dose-response-stimulation model was used for calculating the EC50 of T cell clones. ns not significant, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$.

Declarations

Acknowledgments: We are grateful to 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. Thank Kevin Clark and Sally-Ann Clark from WIMM flow cytometry facility for their help with cell sorting.

This work is supported by UK Medical Research Council (T.D, G.O, Y.P, J.C); Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (CIFMS), China (grant number: 2018-I2M-2-002) (T.D, Y.P, X.Y, G.L, D.D, J.K.K, G.O, G.S, J.M, A.M, A.T); National Institutes of Health, National Key R&D Program of China (2020YFE0202400) (T.D); China Scholarship Council (Z.Y, G.L, C.J, C.L). The study is also funded by the NIHR Oxford Biomedical Research Centre (L.P.H, G.O, J.C.K, A.J.M), Senior Investigator Award (G.O) and Clinical Research Network (G.O); Schmidt Futures (G.R.S); NIHR (SF and Cov19-RECPLAS) and WT109965MA (PK). The McKeating laboratory is funded by a Wellcome Investigator Award (IA) 200838/Z/16/Z, UK Medical Research Council (MRC) project grant MR/R022011/1 and Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Science (CIFMS), China (grant number: 2018-I2M-2-002). Initial funding for the VSF was provided by the Oxford Biomedical Research Centre and Cancer Research UK.

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 T cell experiments; Y.P, D.D, X.Y, G.L, Z.Y, and J.C performed all 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 Class I pentamers; J.C.K, A.J.M, L.P.H, 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, W.D processed clinical samples; R.A.F, C.W, P.S, A.K, C.R-G, R.B-R, I.N, R.A.W, O.T, C.A.T, P.K.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.R.S, and J.M provided technical assistance and critical reagents; Y.P, S.L.F, F.P and G.L analysed data; T.D, J.C.K and Y.P supervised data analysis, T.D , Y.P, S.L.F wrote the original draft. J.C.K, G.O, A.M, G.L.S, H.S, P.K, B.C, P.B and S.L.F reviewed and edited the manuscript and figures.

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials.

References

1. Yewdell, J.W. Confronting complexity: real-world immunodominance in antiviral CD8+ T cell responses. Immunity 25, 533-543 (2006). 2. 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). 3. 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). 4. 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). 5. 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 U S A 117, 24384-24391 (2020). 6. Schulien, I. et al. Characterization of pre-existing and induced SARS-CoV-2-specific CD8(+) T cells. Nat Med 27, 78-85 (2021). 7. 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). 8. de Silva, T.I. et al. The impact of viral mutations on recognition by SARS-CoV-2 specific T-cells. bioRxiv (2021). 9. 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). 10. 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). 11. Sekine, T. et al. Robust T Cell Immunity in Convalescent Individuals with Asymptomatic or Mild COVID-19. Cell 183, 158-168 e114 (2020). 12. 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). 13. 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). 14. et al. A blood atlas of COVID-19 defines hallmarks of disease severity and specificity. medRxiv (2021). 15. 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). 16. 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). 17. Peng, Y. et al. Boosted Influenza-Specific T Cell Responses after H5N1 Pandemic Live Attenuated Influenza Virus Vaccination. Front Immunol 6, 287 (2015). 18. 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). 19. 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). 20. Akthar, S. et al. Matrikines are key regulators in modulating the amplitude of lung inflammation in acute pulmonary infection. Nat Commun 6, 8423 (2015). 21. 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 U S A 109, 13353-13358 (2012). 22. 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). 23. Singh, Y. et al. Differential effect of DJ-1/PARK7 on development of natural and induced regulatory T cells. Scientific Reports 5, 17723 (2015). 24. Danileviciute, E. et al. PARK7/DJ-1 promotes pyruvate dehydrogenase activity and maintains Treg homeostasis. bioRxiv (2019). 25. Johansson, C. & Kirsebom, F.C.M. Neutrophils in respiratory viral infections. Mucosal Immunol (2021). 26. 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). 27. Abd Hamid, M. et al. Self-Maintaining CD103^{+} Cancer-Specific T Cells Are Highly Energetic with Rapid Cytotoxic and Effector Responses. Cancer Immunology Research 8, 203-216 (2020). 28. Zhang, J. et al. A systemic and molecular study of subcellular localization of SARS-CoV-2 proteins. Signal Transduction and Targeted Therapy 5, 269 (2020). 29. 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 U S A 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. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc 9, 171-181 (2014). 35. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21 (2013). 36. 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). 37. Stuart, T. et al. Comprehensive Integration of Single-Cell Data. Cell 177, 1888-1902 e1821 (2019). 38. Wickham, H. ggplot2 : Elegant Graphics for Data Analysis. Use R!,. 2nd ed. Cham:

Springer International Publishing : Imprint: Springer,; 2016. pp. 1 online resource (XVI, 260 pages 232 illustrations, 140 illustrations in color. 39. Bolotin, D.A. et al. Antigen receptor repertoire profiling from RNA-seq data. Nature Biotechnology 35, 908-911 (2017). 40. Bolotin, D.A. et al. MiXCR: software for comprehensive adaptive immunity profiling. Nature Methods 12, 380-381 (2015). 41. Gu, Z., Gu, L., Eils, R., Schlesner, M. & Brors, B. circlize Implements and enhances circular visualization in R. Bioinformatics 30, 2811-2812 (2014). 42. van der Loo, M. The stringdist Package for Approximate String Matching. R Journal (2014). 43. Brunson, J.C. ggalluvial: Layered Grammar for Alluvial Plots. Journal of Open Source Software (2020). 44. Zheng, G.X. et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun 8, 14049 (2017). 45. Huang, X. & Huang, Y. Cellsnp-lite: an efficient tool for genotyping single cells. Bioinformatics (2021). 46. 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). 47. Amezquita, R.A. et al. Orchestrating single-cell analysis with Bioconductor. Nat Methods 17, 137-145 (2020).

Table 1

Table 1. Groups defined by shared TRBV gene usage and CDR3β sequence between bulk TCR sequencing from T cell clones and single cell TCR sequencing from *ex vivo* T cells.

Group	CDR3β	TRBV	TRBJ	Functional avidity
1	CAISEPGTSGGAILDTQYF	TRBV10-3	TRBJ2-3	Low
2	CASGPATSAEQETQYF	TRBV12-5	TRBJ2-5	High
3	CASSILQGLGGSNQPQHF	TRBV19	TRBJ1-5	Low
4	CASSVLPGPPRGEQFF	TRBV2	TRBJ2-1	High
5	CSAQVGGNYNSPLHF	TRBV20-1	TRBJ1-6	High
6	CATSDLVTSGDEQFF	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	CASSPLTGSAERKETQYF	TRBV27	TRBJ2-5	High
11	CASSPLVGERFRKETQYF	TRBV27	TRBJ2-5	Low
12	CASSSLLAGGFYEQFF	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	CASSSYPGLAPVQETQYF	TRBV5-1	TRBJ2-5	High
18	CASSYLPAGSSYNSPLHF	TRBV6-3	TRBJ1-6	High

Figures

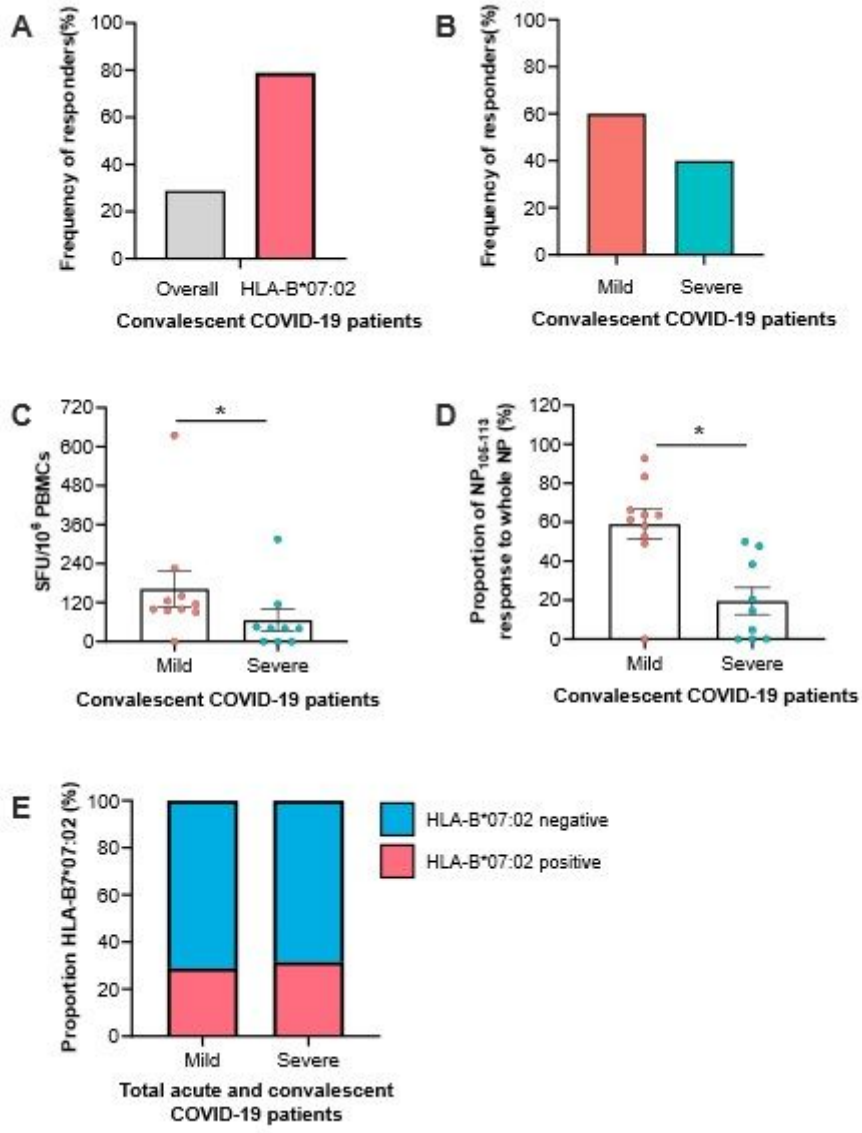

Figure 1

Frequency and magnitude of response to NP105-113-B*07:02 epitope in COVID-19 patients. (A) Frequency of convalescent COVID-19 patients (N=52 overall and HLA-B*07:02 positive, N=19) with T cells responding to NP105-113-B*07:02 epitope stimulation. (B) Frequency of responders (N=15) with mild or severe convalescent COVID-19 disease. (C) Comparison of magnitude of response to NP105-113 epitope between HLA-B*07:02 positive convalescent COVID-19 patients (N=10 mild, N=9 severe, * P<0.05). (D) Proportion of NP105-113-specific response to overall nucleoprotein response (N=10 mild, N=9 severe, * P<0.05). (E) Proportion of HLA-B*07:02 individuals compared to combined total acute and convalescent COVID-19 patients (N=77 acute, N=52 convalescent).

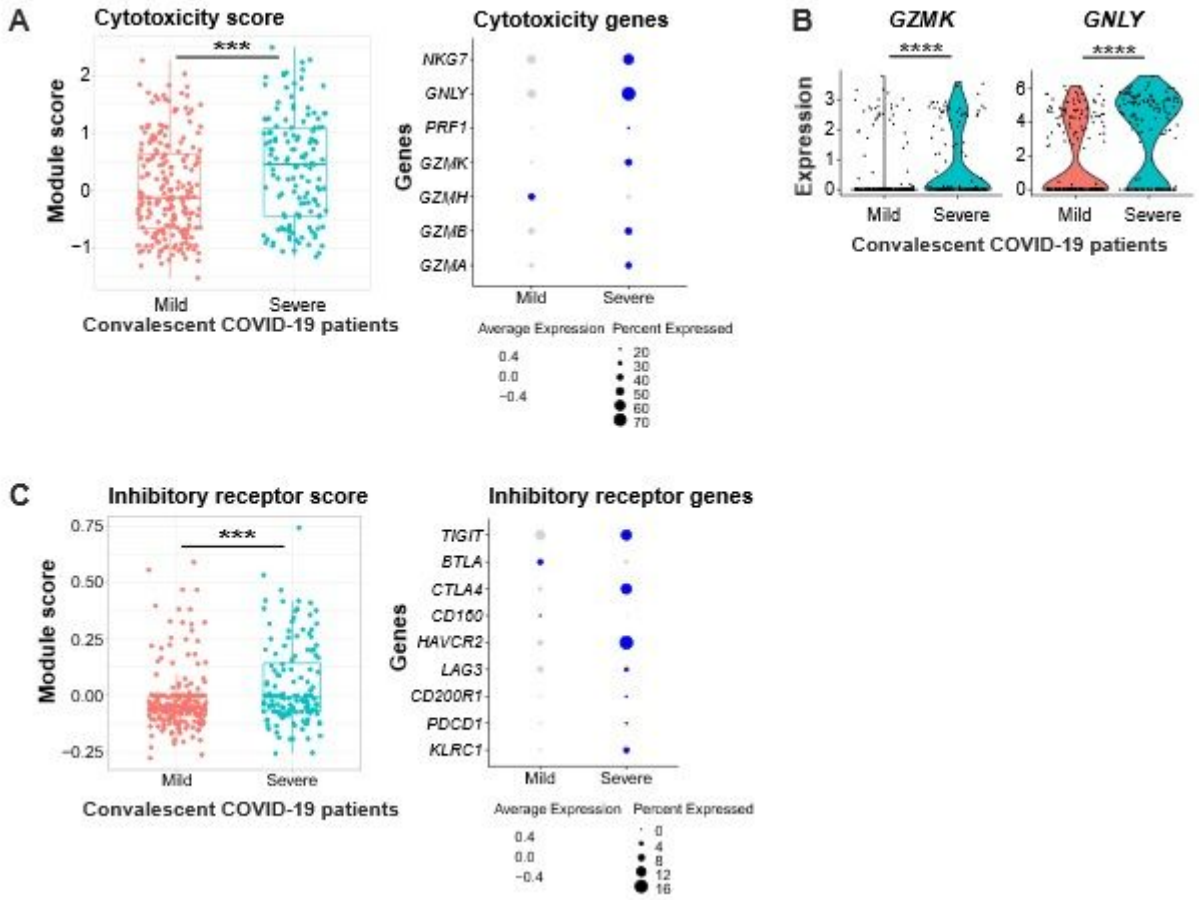

Figure 2

Characterisation of response to NP105-113-B*07:02 epitope in convalescent HLA-B*07:02 positive mild and severe COVID-19 patients. (A) Gene sets were scored based on single cell gene expression from SmartSeq2 RNA-Seq dataset comprised of two mild and two severe convalescent HLA-B7*07:02 positive COVID-19 patients (208 cells from mild cases, 140 cells from severe cases). Scores for cytotoxic gene expression are shown in box plot (*** P<0.001). Individual genes in the cytotoxic gene set are shown on the right panel. (B) Violin plots showing specific expression of cytotoxic genes (**** P<0.0001). (C) Box plot showing scores for inhibitory receptor gene set (*** P<0.001), right panel shows individual genes in gene module.

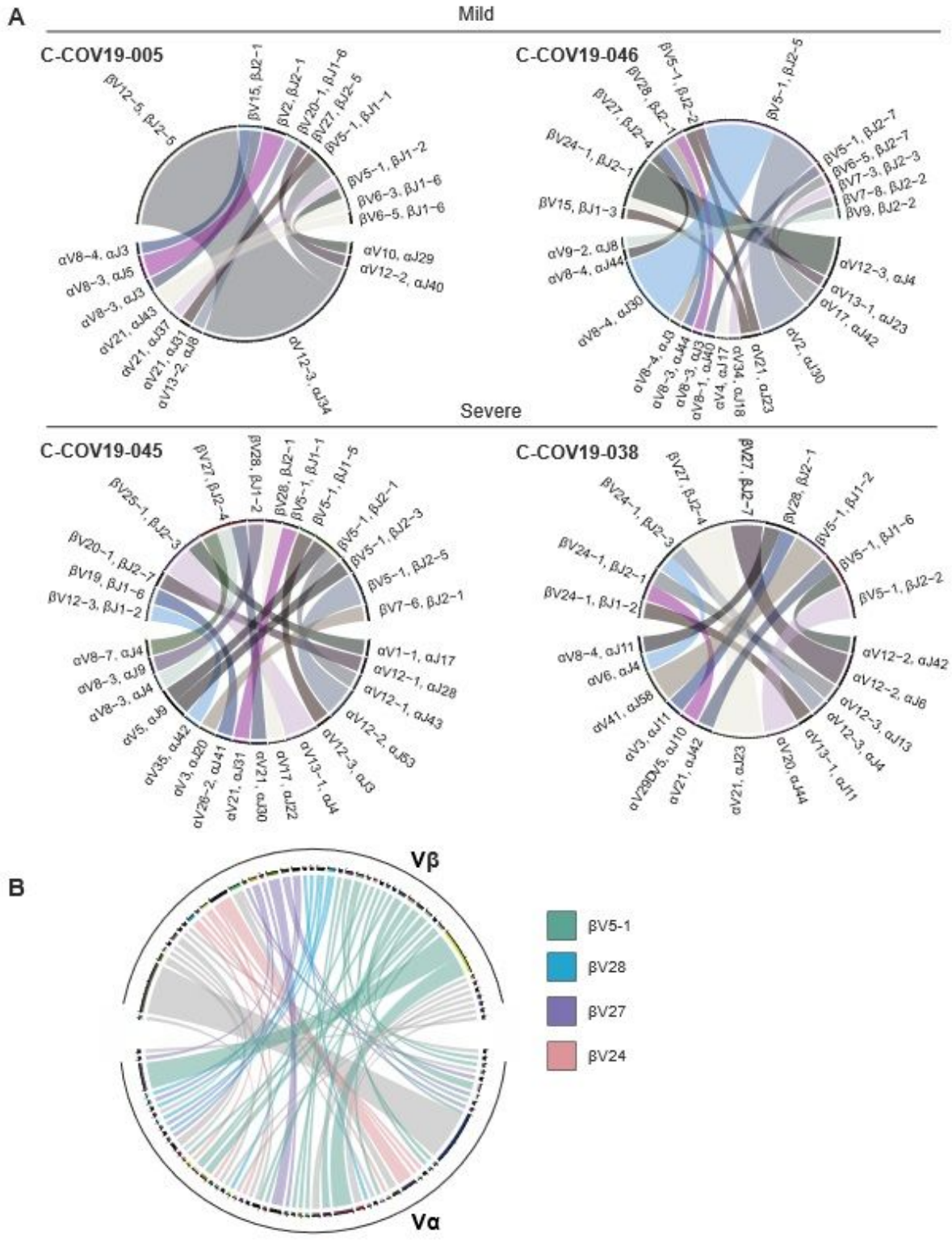

Figure 3

Paired αβ TCR repertoire of NP105-113-B*07:02 specific T cells from convalescent COVID-19 patients. (A) Circos plots for each patient depicting αβ 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 Vα and Vβ). Each line represents a unique clonotype. Clonotypes that have

dominant Vβ gene usage (TRBV5-1, TRBV28, TRBV27 and TRBV24) are highlighted; all others are show in grey.

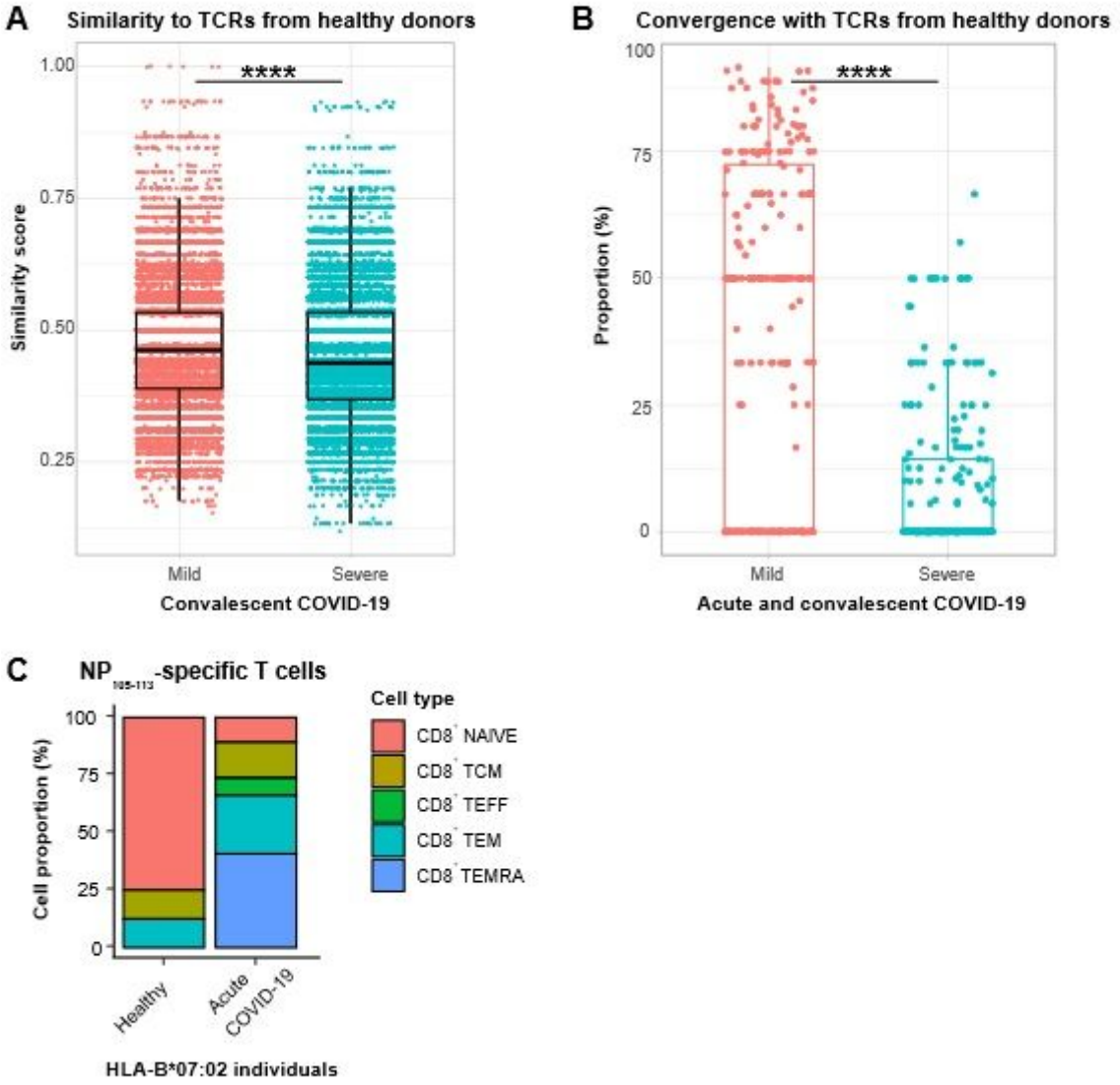

Figure 4

Comparison and characterisation of NP105-113-B*07:02-specific TCRs from acute and convalescent COVID-19. (A) Similarity scores from pairwise comparisons between TCRs from pre-pandemic individuals (237 TCRs) and 85 unique clonotypes from convalescent COVID-19 patients (mild v severe, **** P<0.0001). (B) Proportion of acute and convalescent TCRs from mild and severe COVID-19 patients found in the same GLIPH2 convergence groups as TCRs from 12 healthy donors (738 TCRs from 12 mild patients and 133 TCRs from 7 severe patients in 264 convergence groups, mild v severe **** P<0.0001). 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-B7*07:02 positive donor (8 cells) and HLA-B*07:02 positive COVID-19 patients at acute

stage (130 cells from 17 COVID-19 patients). TCM: T central memory, TEFF: T effector, TEM: T effector memory, TEMRA: T effector memory re-expressing CD45RA.

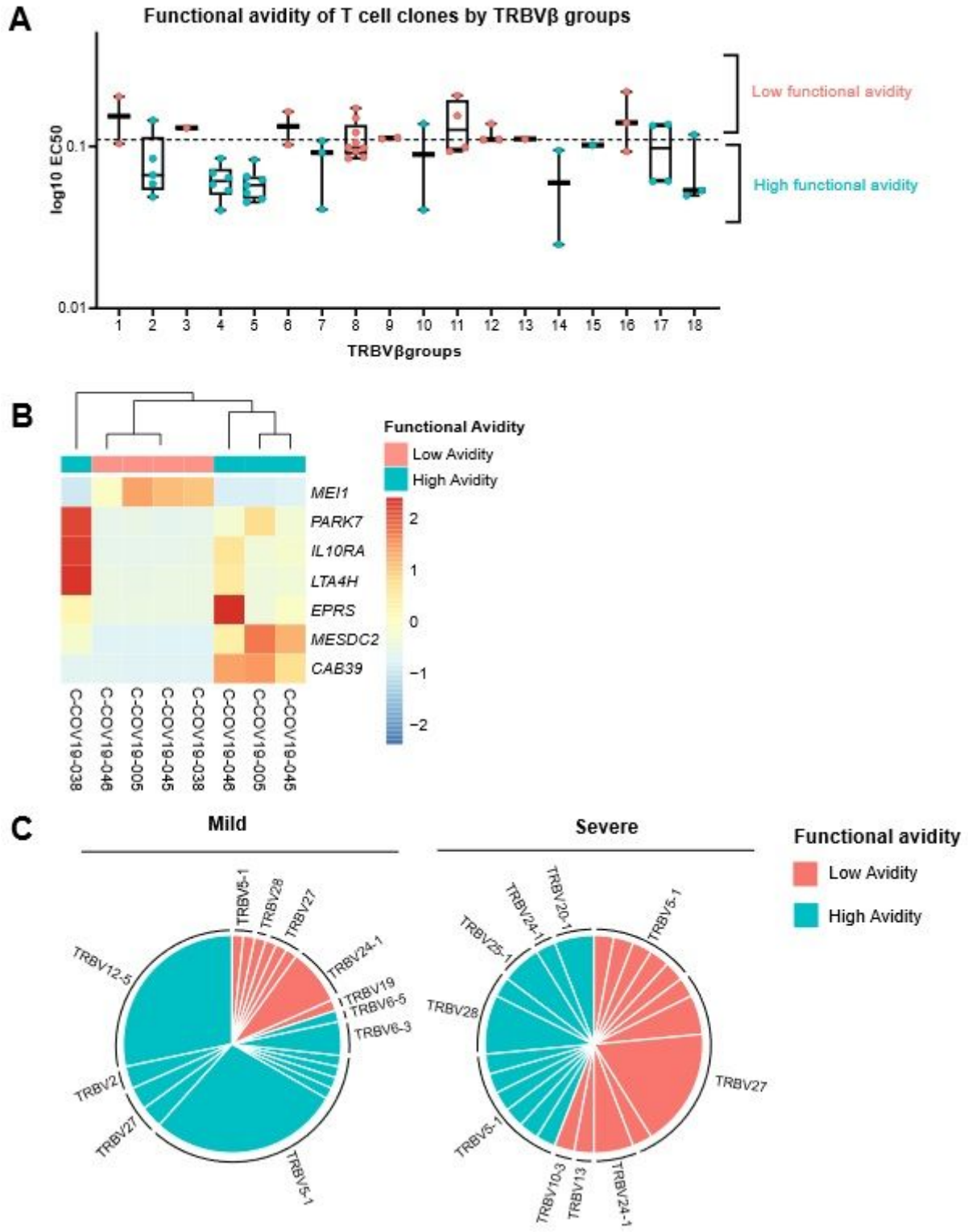

Figure 5

Functional avidity and clonotype expansion of NP105-113-B*07:02-specific T cells. (A) Functional avidity of T cell clones by TRBV groups. 60 NP105-113-B*07:02-specific T cell clones were derived from four convalescent COVID-19 patients and functional avidity (EC50) was measured by IFN-γ ELISPOT assay.

Based on their TRBV gene usage and CDR3β sequences, T cell clones and single cell counterparts were sorted into 18 distinct groups, and further divided into high or low functional avidity groups (cut-off EC50 0.11). (B) Heat map showing differential gene expression comparing “pseudobulk” high and low functional avidity single cells (88 high avidity cells and 52 low avidity cells, genes shown have adjusted P<0.05). (C) Comparison of functional avidity and expansion of TCR clonotypes (defined as TRBV gene usage and CDR3β sequence in each patient) in mild and severe convalescent COVID-19 patients (N=4).

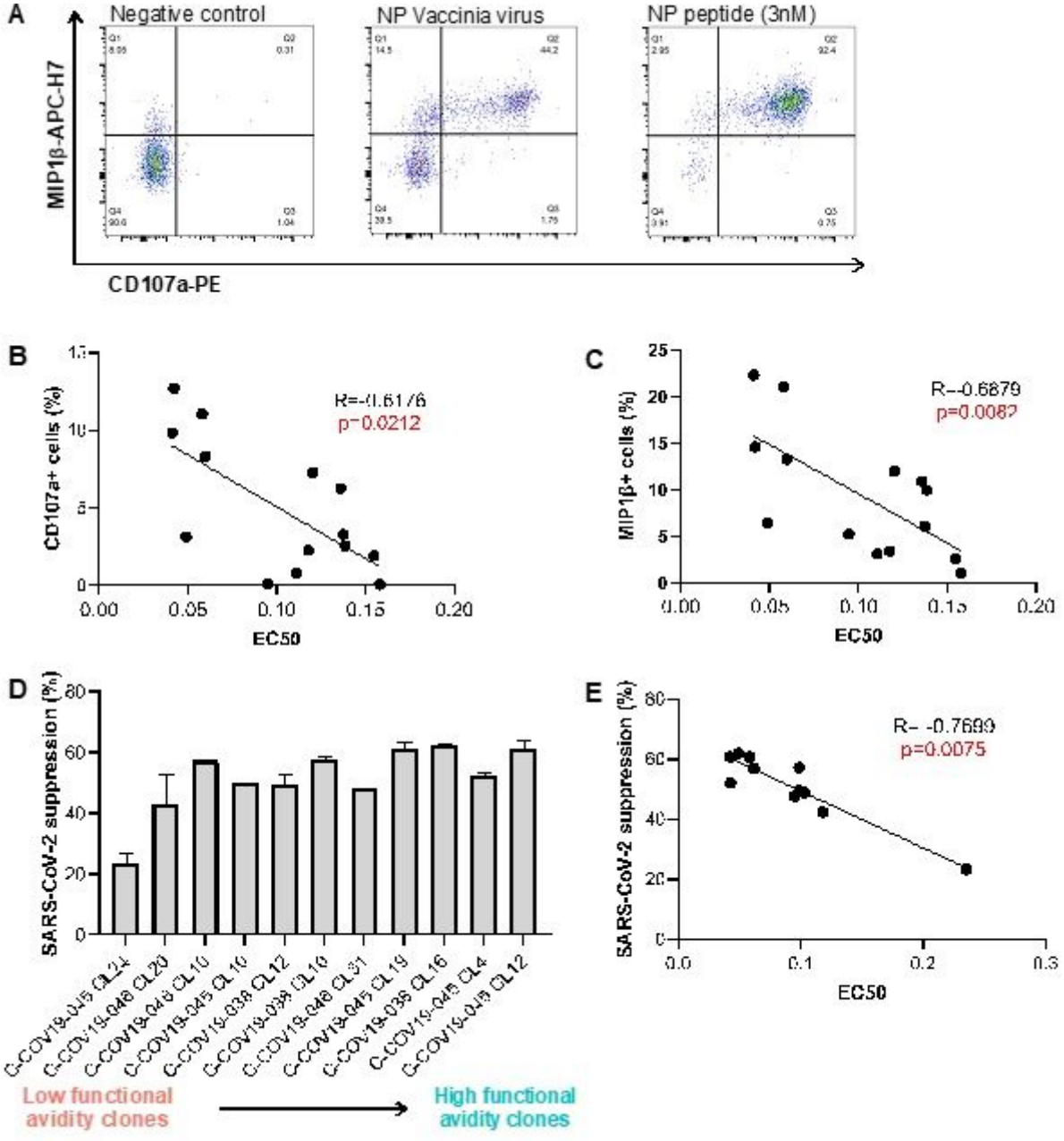

Figure 6

Correlation between functional avidity and anti-viral efficacy in T cell clones. (A) Representative ICS flow cytometry plots measuring MIP1β and CD107a expression on T cell clones incubated with Vaccinia virus encoding nucleoprotein or peptide-loaded (3nM 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, P<0.05). (C) Correlation plot between MIP1β production of T cell clones incubated with NP-expressing Vaccinia virus and their respective EC50 values (N=14, P<0.01). (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, P<0.01).

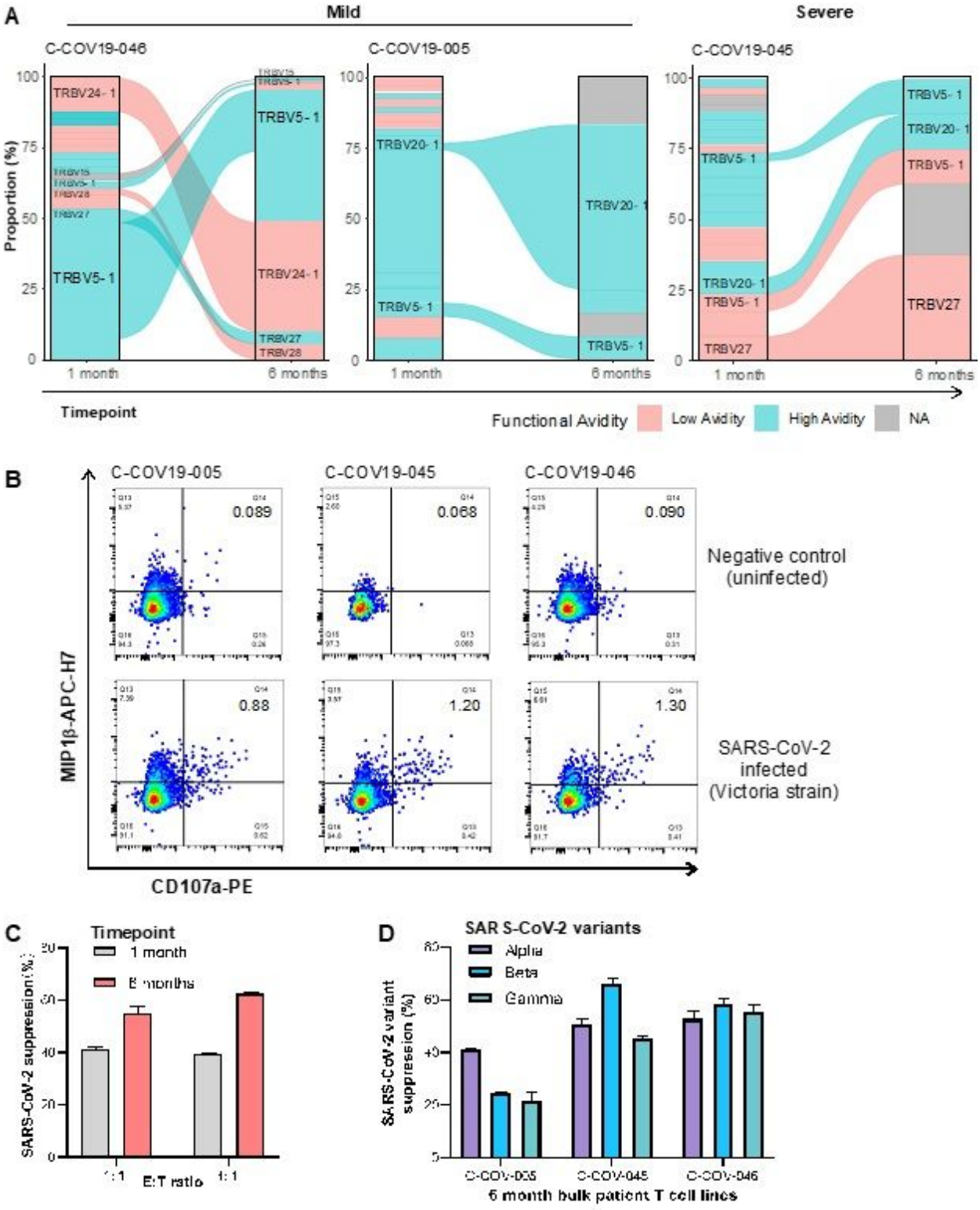

Figure 7

Characterisation of NP105-113-B*07:02 specific T cell responses at six months convalescence. (A) TCR repertoires of three patients at one month and six months convalescence. TRBV gene usage of common and expanded TCR clonotypes (defined as TRBV and TRBJ gene usage) are labelled for clarity. TCR clonotypes coloured pink are low functional avidity, blue depict high functional avidity; clonotypes coloured grey do not have similar TCRs to T cell clones. C-COV19-46 six months cells were sequenced by 10x single cell sequencing, C-COV19-005 and C-C0V19-045 were bulk TCR sequencing. (B) Representative ICS flow cytometry plots measuring MIP1β and CD107a expression on bulk NP105-113-specific T cell lines incubated with SARS-CoV-2 (Victoria strain) infected autologous B cell lines. (C) Inhibition of SARS-CoV-2 viral replication (Victoria strain) by C-COV19-046 bulk NP105-113-specific T cell lines from one month (grey bars) and six months (red bars) convalescent samples. Single representative experiment shown (error bars indicate technical replicates). E:T – effector: target ratio. (D) Anti-viral activity of NP105-113-specific bulk T cells from six months convalescence against SARS-CoV-2 variants of concern: Alpha (purple bars), Beta (blue bars) and Gamma (green bars). Single representative experiment shown (error bars indicate technical replicates).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- NIA32469Asuppfigs.docx
- NIA32469Asupptable2.xlsx
- NIA32469Asupptable3.xlsx
- NIA32469Asupptable4.xlsx
- NIA32469Asupptable5.xlsx