

Title: Large clones of pre-existing T cells drive early immunity against SARS-COV-2 and LCMV infection.

Martina Milighetti^{1†}, Yanchun Peng^{2,3†}, Cedric Tan^{5†}, Michal Mark^{6†}, Gayathri Nageswaran¹, Suzanne Byrne¹, Tahel Ronel¹, Tom Peacock¹, Andreas Mayer¹, Aneesh Chandran¹, Joshua Rosenheim¹, Matthew Whelan¹, Xuan Yao³, Guihai Liu³, Suet Ling Felce³, Tao Dong^{2,3}, Alexander J. Mentzer⁴, Julian C. Knight^{3,4}, Francois Balloux⁵, Erez Greenstein⁶, Shlomit Reich-Zeliger⁶, Corrina Pade⁷, Joseph M. Gibbons⁷, Amanda Semper⁷, Tim Brooks⁸, Ashley Otter⁸, Daniel M Altmann⁹, Rosemary J Boyton^{10,11}, Mala K Maini¹, Aine McKnight⁷, Charlotte Manisty¹², Thomas A. Treibel¹², James C. Moon¹², COVIDsortium Investigators[‡], Mahdad Noursadeghi¹, Benny Chain^{1*}

Affiliations:

1. Division of Infection and Immunity, University College London; London WC1E 6BT, UK.
2. MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford; Oxford, UK.
3. Chinese Academy of Medical Science (CAMS) Oxford Institute (COI), University of Oxford; Oxford, U.K.
- 15 4. Wellcome Centre for Human Genetics; University of Oxford; Oxford, U.K.
5. UCL Genetics Institute, University College London; London WC1E 6BT, UK.
6. Department of Systems Immunology, Weizmann Institute of Science; Rehovot 7610001, Israel.
7. Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London; London E1 4NS, UK.
- 20 8. UK Health Security Agency; Porton Down, Salisbury, SP4 0JG, UK.
9. Department of Immunology and Inflammation, Imperial College London; London SW7 2BX, UK.
10. Department of Infectious Disease, Imperial College London; London, W12 0NN, UK.
11. Lung Division, Royal Brompton Hospital, Guy's and St Thomas' NHS Foundation Trust; London, UK.
- 25 12. Institute of Cardiovascular Sciences, University College London; London WC1E 6BT, UK.

† + these authors contributed equally to the study

‡ Full list of the authors and affiliations is given in the Supplementary Materials

- Corresponding author b.chain@ucl.ac.uk

Abstract: We analyzed the dynamics of the earliest T cell response to SARS-COV-2. A wave of TCRs strongly but transiently expand during infection, frequently peaking the same week as the first positive PCR test. These expanding TCR CDR3s were enriched for sequences functionally annotated as SARS-COV-2 specific. Most epitopes recognized by the expanding TCRs were highly conserved between SARS-COV-2 strains, but not with circulating human coronaviruses. Many expanding CDR3s were also present at high precursor frequency in pre-pandemic TCR repertoires. A similar set of early response TCRs specific for lymphocytic choriomeningitis virus epitopes were also found at high frequency in the pre-infection naïve repertoire. High frequency naïve precursors may allow the T cell response to respond rapidly during the crucial early phases of acute viral infection.

One-Sentence Summary: High frequency naïve precursors underly the rapid T cell response during the crucial early phases of acute viral infection.

Main Text:

T cell responses to vaccination (1) and acute infection (2) are often fast, with the magnitude of the response peaking within one to two weeks of exposure, and substantially preceding maximum antibody responses. T cells may therefore combine with innate immunity to control microbial growth early in infection, before humoral immunity is fully developed. The mechanisms which underly this early clonal expansion remain poorly defined. In this study, we exploit a granular longitudinal sampling of peripheral blood cells which were collected during, and in some cases preceding acute infection with SARS-CoV-2, during the first few weeks of the pandemic (3). The dynamic changes in the T cell receptor repertoire which are observed during this period reveal mechanisms that allow the speed of the T cell adaptive response to compete with that of innate immunity and viral replication during the crucial early phases of this acute viral infection.

There is an increasing interest in the role of T cells in providing protection against SARS-CoV-2 infection and disease (4). HLA-peptide binding and in vitro peptide-driven expansion have provided a comprehensive catalogue of T cell epitopes (5, 6), which span most of the viral open reading frames (ORFs). Single cell protein and transcriptomic analyses have provided complementary information on memory and effector T cell function (7, 8). There have also been many attempts to link T cell responses to clinical outcome, although distinguishing cause and effect in these observational studies is challenging, especially because SARS-CoV-2 disease is often associated with lymphopenia. Nevertheless, there is some evidence that T cell responses may be associated with better clinical outcomes (9–11) and this has fueled the development of more T-cell centric vaccines, especially in individuals with impaired humoral immunity(12).

Despite the wealth of published studies on SARS-CoV-2, however, dynamic analysis of immune responses in the early phases of infection are still rare(13). Indeed, detailed dynamics of the earliest responses are rare in any human infection. Experimental infection studies (e.g. to influenza) have provided some important data in this regard, but prior exposure is often a confounder in this setting (14). We have previously reported on the COVIDSortium (15), a longitudinal study of London-based health care workers, established very early in the first wave of the COVID pandemic, which provides a detailed granular view of the peripheral blood compartment during the first few weeks of infection (3). We have used this sample set to demonstrate a remarkably fast CD4 and CD8 T cell response following SARS-CoV-2 infection, which parallels the innate interferon response, and substantially precedes humoral immunity (16). In the present study, we investigate in more detail this early T cell response, using robust quantitative T cell repertoire sequencing to chart the evolution of the anti-viral response at the level of the clonotype.

Global TCRseq provides an unbiased survey of the T cell response, which is complementary to more antigen-specific methodologies which depend on predefined targets such as peptides or HLA multimers. We identify a wave of early T cells, which are enriched for sequences independently annotated as SARS-CoV-2-specific. Many of the CDR3 sequences are found in the pre-pandemic repertoires of healthy individuals, and we leverage this information to develop a statistical framework to infer that they are present at high clonal frequency in pre-pandemic repertoires. Remarkably, we extend these observations to similar early expanding epitope-specific T cell receptors (TCRs) in LCMV infection, and demonstrate that many of these sequences can be found at high abundance in naïve repertoires prior to and independently of exposure to lymphocytic choriomeningitis virus (LCMV). The ability to rapidly recruit high abundance TCRs commonly found in the naïve TCR repertoire may be a common strategy to cope with the response to novel microbial pathogens.

Results

We sequenced the T cell repertoire of whole blood RNA samples collected at different time points (see below) from 41 health care workers (HCW) who tested PCR+ for SARS-COV-2 infection, and 6 HCW who remained PCR and seronegative throughout the study (sample collection summarized in Fig S2). The samples were collected at weeks 0 - 4 (acute) and week 12-14 (convalescent). The median number of samples collected for each individual was 4, ranging from 1 to 6. The majority of infections registered a PCR+ test at the first sample, but in 12 individuals, samples predating the first PCR+ sample were obtained. The first 3-4 weeks of sample collection coincided with the first wave of the SARS-COV-2 pandemic in London, and no additional cases of PCR+ confirmed infection or seroconversion were observed in the subsequent 2-3 months of collection. We sequenced a total of 14.6 million TCR alpha and TCR beta genes, with a median 73,000 TCR alpha and 94,000 TCR beta sequences per sample (Fig S3A). The number of alpha and beta sequences per sample were highly correlated (Spearman correlation = 0.93, Fig S3B), providing additional confidence in the quantitative robustness of the pipeline.

We first measured the richness (defined as number of distinct sequences, divided by total number of sequences in sample) and Shannon diversity (which incorporates a measure of the frequency distribution of the sequences) of the repertoires at different time points in both PCR+ and PCR- individuals. For PCR+ individuals we plotted the values relative to the week at which the individual first became PCR+ (Fig S4). However, there were no detectable effects of infection on the global parameters of the TCR repertoire.

We therefore focused on identifying individual TCRs which changed significantly in abundance within the first 5 weeks of the study (see Materials and Methods and (15) for further details). In some pairwise comparisons a clear population of expanded TCRs was observed (Fig 1A). In others, especially where the first sample available was already PCR+ , a population of contracting TCRs was observed (Fig S5A). In these cases, we made the assumption that we had missed the expansion phase, but observed the contraction phase of the initial T cell response. We therefore combined all up and downregulated TCRs, which we refer to as expanded TCRs, and removed MAIT cells (defined as TRAV1-2 paired with TRAJ12, TRAJ20 or TRAJ33) and iNKT cells (identified as TRAV10 paired with TRAJ18) from further analysis, leaving 4075 expanded TCR alpha sequences and 5458 expanded TCR beta sequences. In control uninfected individuals there were few expanding or contracting TCRs (Fig S5B). The richness and diversity of the set of expanded TCRs increased at the time of first PCR+ in infected individuals (p-value < 0.05 for all time points compared to pre-infection, Wilcoxon signed-rank test), but showed no consistent dynamics in uninfected individuals (Fig 1B). Note that the time axis was rescaled relative to the week at which that individual became PCR+ (renamed as week 0).

We plotted the combined abundances of all expanded TCRs for each individual (an estimate of total clonal expansion) as a function of time (Fig 1C). The magnitude of the response varied considerably between individuals, but showed a common dynamic. The response was already strong at week 0 and fell by week 14. In some individuals responses could be detected even in the week prior to becoming PCR+. The magnitude of the response of the “expanded” TCRs in the PCR- seronegative controls showed no evidence of any common dynamic pattern (Fig 1D). For comparison we show the dynamics of the anti-spike (Fig S6A) and anti-nuclear protein (Fig S6B) antibody titres, which rose more gradually and slowly over the first 14 weeks, as described in detail previously (34).

5

An example of the time course of an individual HCW, for whom we had samples pre as well as post infection, shows more clearly the rapid expansion of individual TCR alpha and TCR beta sequences, which were mostly not detected pre-infection, because of sampling only a very small proportion of the repertoire, but maximal at either week 0 (week first PCR+) or week 1 (Fig 1E). Although the magnitude of the T cell response returns towards baseline by week 14, the expanded TCRs in this individual remain significantly above their starting abundance (Fig S7). For comparison, the timecourse of TCRs in an individual who did not become PCR+ show no clear pattern over the weeks of sampling (Fig 1F). The individual time courses for TCRs for all HCW are shown in Fig S8.

10 In summary, analysis of the dynamics of the TCR repertoire over the first 4 weeks identified a set of TCRs whose abundance increased rapidly and transitorily around or shortly after the time of SARS-CoV-2 detection.

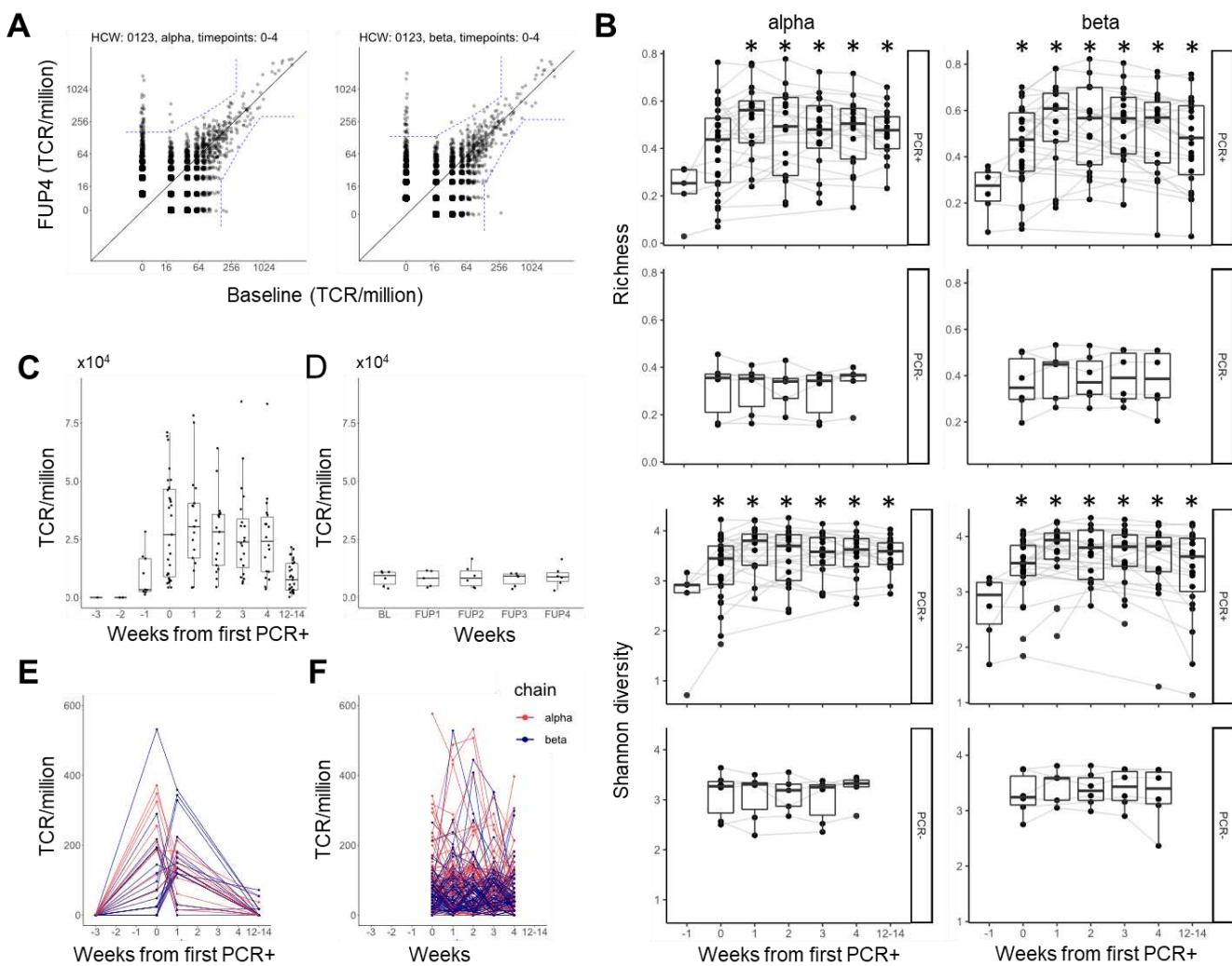


Fig 1. An early transient wave of TCR expansion associated with infection with S2.

15

A. An example of a pairwise comparison between two time points, showing TCRs expanding between baseline and follow up week 4 (FUP4) repertoires. The individual (ID 123) became PCR+ at follow up week 1 (FUP1). Each point is an individual TCR sequence, plotting its abundance at FUP4 versus its abundance at Baseline. All abundances are normalized to number of TCRs per million. The dashed blue line indicates the significance threshold calculated as described in M&M. All TCRs which fall outside the dashed line are considered as expanded (contracted).

B. The richness and Shannon diversity of the set of expanded (contracted) TCRs at each time point, subsampled to similar dataset size. For PCR+ individuals, the x-axis is rescaled relative to the week at which they first became PCR+ (this is week 0). For PCR- individuals the weeks correspond to baseline and subsequent follow-ups at weeks 1-4. * p-value < 0.05, Wilcox test compared to week -1.

5 C. The sum of the expanded TCRs at each timepoint, for each infected individual. The timepoints have been recalibrated relative to the week of the first PCR+ test (this is defined as week 0). The boxplots show median, interquartile (box) and 95% (whiskers) range.

D. As for C, but for controls who did not become PCR+.

10 E. The dynamics of individual TCRs (TCRalpha in red, TCRbeta in blue) in one individual (ID 101).

F. As for E but for a PCR- (ID 17). The dynamics for each individual HCW are shown in supplementary data.

TCRs expanding early are enriched for sequences associated with SARS-COV-2 recognition.

We collected a set of 7694 (7632 unique across all sets, Fig 2A) functionally annotated TCRs, combining from a public TCR database VDJDb (35), a published study (36) and a set of TCRs we identified by 15 tetramer sorting or in vitro peptide expansions (23). The functional affinity of the TCRs varied over a broad range, but included several high affinity clones (Fig S9). 239 (141 unique) CDR3s from the expanded TCR set matched with CDR3 sequences in the annotated TCR sequences (3.1%, Fig 2B). The majority of matches were with alpha chains. Same size random sets of non-expanded TCRs from the HCW repertoires matched an average of 4.2 (0.06%) annotated TCRs, significantly (more than 10 fold) 20 less than the number of annotated sequences in the expanded set ($P < 0.0001$, Fisher's exact test). There were only 7 matches of SARS-COV-2 annotated CDR3 with the set of expanded TCRs from the control (PCR-) repertoires.

We examined whether the expanded sets were also enriched for CMV annotated or EBV annotated TCR, 25 since these viruses are prevalent in the general population (Fig 2C). The enrichment for SARS-COV-2 annotated TCRs was significantly higher than for CMV or EBV. As a comparison, we looked for matches with HIV annotated TCRs, since this virus is not likely to be prevalent in the general population. The proportion of matches with the HIV set was smaller than CMV or EBV, but the lack of annotated HIV-specific TCRalpha sequences in the database complicates interpretation.

Sharing of an identical CDR3 sequence between SARS-COV-2 annotated TCRs and the expanded set of 30 TCRs identified in this study does not necessarily guarantee that the antigen target of the CDR3s is the same. However, we observed that the expanded CDR3 in our data was associated with the same TCR V gene as reported in the annotated set (Fig 2D, left). In addition, the HLA of the individuals in whom we observed expanded matched CDR3s was associated with the HLA restriction reported for the epitope 35 reported for each annotated TCR (Fig 2D, right). The sharing of CDR3 sequence, V gene and HLA restriction are together suggestive of shared epitope recognition.

The distribution of annotated expanded TCRs among the PCR+ HCW cohort, and the associated 40 metadata available for each epitope are shown in Fig 2E. Almost every PCR+ individual contained at least one annotated expanded TCR, and many contained several. The annotated matched TCRs included both CD4+ and CD8+ cells, were restricted by a variety of HLA alleles and recognized epitopes from a variety of structural and non-structural viral proteins. The time course of the annotated TCRs broadly matched that of the total set of expanded TCRs, with the majority of TCRs peaking the week of the first PCR+ test (Fig S10).

A number of expanded TCRs (including some expanded annotated TCRs, Fig 2D) were found in more than one individual. Since TCRs with similar CDR3 sequence often recognize the same antigen (37-39)

5

we looked for sharing of similar, as well as identical expanded TCRs between individuals. Similarity was measured using a triplet kernel as described previously (40) and in M&M. We observed that a substantial proportion of the expanded TCRs formed inter-individual clusters of similar sequence, indicating that similar TCRs were frequently expanded in different individuals following SARS-COV-2 infection (Fig 2F). The proportion of control (non-expanded) TCRs forming clusters was much less (Fig S11). Some of the clustered expanded TCR sequences were also functionally annotated, and where more than one annotated TCR was found within a cluster, the target antigen was generally the same (Fig S12).

10 Taken together, the data presented in Fig 2 support the hypothesis that the early wave of expanding TCRs which follow infection with SARS-COV-2 represent a set of virus specific TCRs, many of which share similar or identical sequences in different individuals.

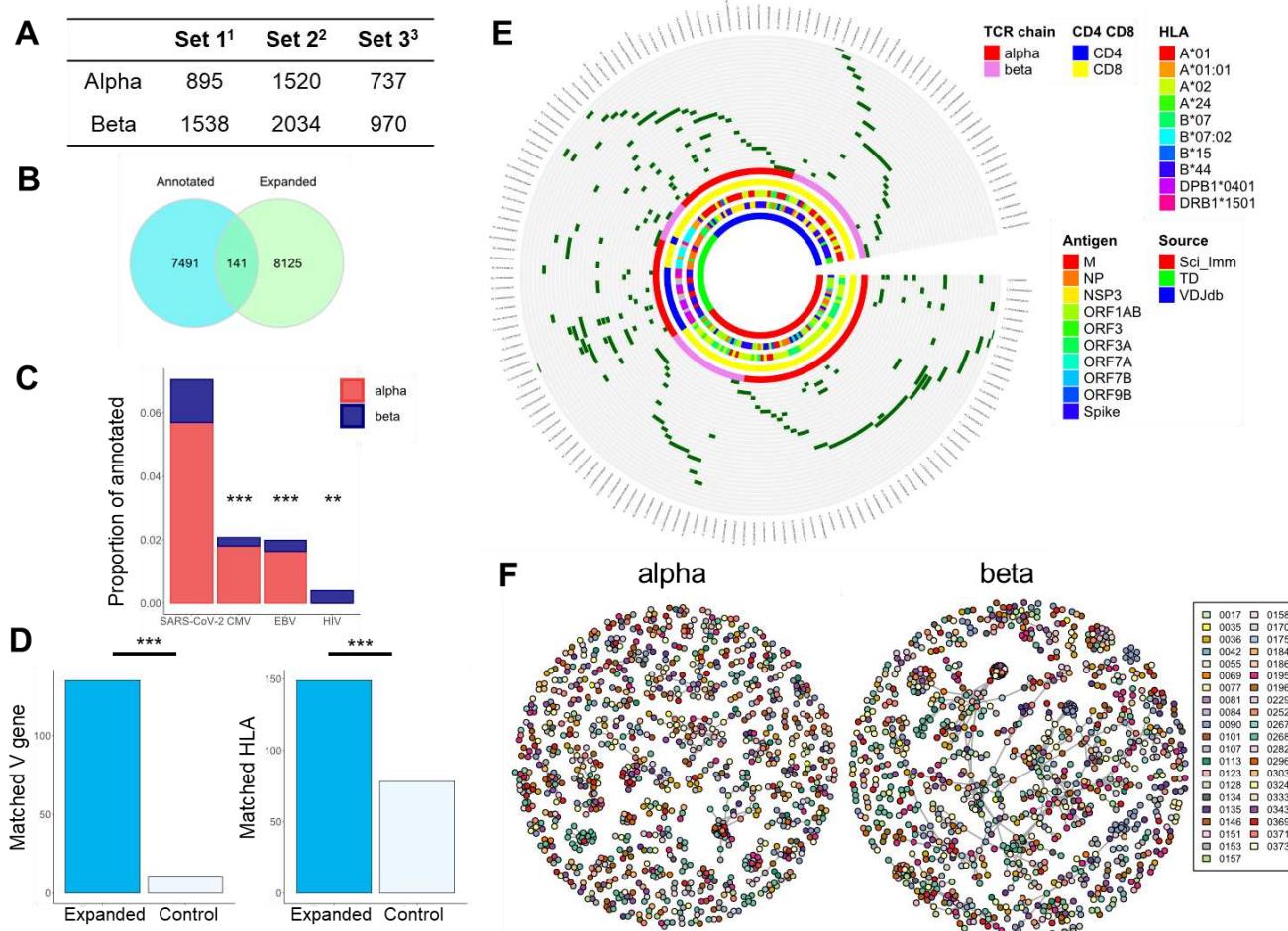


Fig 2. Functional annotation for SARS-COV-2 and sequence similarity characterise the early wave of expanded TCRs in PCR+ individuals.

15

A. We collected a set of TCR sequences which had been functionally annotated for recognition of SARS-COV-2 epitopes. The set was compiled from three data sets (3423) (36) (3423). The set combines paired alpha/beta sequences and unpaired sequences. The Table shows the number of alpha and beta sequences in each set.

20

B. The intersection between the expanded TCRs as defined in Fig 1 (blue), and the SARS-COV-2 annotated TCRs as in panel A (green).

C. The proportion of annotated TCRs for each of four viruses which are observed in the expanded set of TCRs defined in Fig 1. SARS-COV-2 annotated TCRs are found significantly more often than those for CMV, EBV and HIV (**p <0.0001, ** p<0.001, Fisher's exact test).

5 D. Left panel. Blue shading - the number of expanded TCR CDR3s which share the same V gene as the respective identical annotated TCR CDR3. Empty bars – the number of matches after random shuffling of the expanded set with respect to their V gene. Right panel. Blue shading – the number of annotated TCRs whose restricting HLA gene matches at least one allele of the individual in whom the identical expanded CDR3 is detected. Empty bars – the number of matched TCRs after random shuffling of the annotated TCRs with respect to their HLA restriction (**p<0.0001, Fisher's exact test).

10 E. A circus plot showing distribution of the annotated expanded TCRs in the cohort, and the metadata associated with each TCR. Each segment is a unique CDR3. Each circle is a different individual. Dark green segments correspond to an annotated expanded TCR. The inner circles show which CDR3 are alpha or beta; which CDR3 are derived from CD4 or CD8 T cells; the restricting HLA allele for each CDR3; the target antigen recognised by the annotated TCR; and the source of the CDR3.

15 F. A graph representation of the TCR alpha and beta sequence similarity. Each node is a CDR3 sequence, and edges connect all nodes with a string kernel similarity index of greater than 0.76 For TCRalpha and 0.72 TCRbeta. Each individual is shown in a different colour. Only clusters of three or more connected nodes are shown.

20 **Expanding TCRs associated with SARS-COV-2 infection are abundant in healthy pre-pandemic repertoires.**

The peak in TCR expansion observed around the time of PCR+ virus detection suggests a very rapid T cell response. We explored the possibility that the rapid expansion of TCRs arises from sequences which are present at higher than average frequency in the pre-infection repertoire. We developed a statistical framework to address this hypothesis using a large set of published PBMC repertoires (Fig 3A) .

25 We determined the proportion of repertoires from a cohort of 786 individuals collected and sequenced pre-SARS-Cov-2 (referred to here as the Emerson data set (32), which contain each of the expanded CDR3 sequences defined in our study. Since each repertoire from the Emerson set contains only a small sample of the total repertoire in each individual (on average 1.8×10^5 sequences per sample), the inter-individual sharing must reflect a high abundance of the shared TCR in the sample. The average abundance of the TCR in the repertoire can be estimated for all those TCRs which are seen at least twice in the Emerson data set using the Poisson probability distribution. Specifically, the proportion of the individuals which do NOT contain a particular sequence ($p(X = 0)$) is given by

$$Eq\ 1.\ p(X=0) = e^{(-\mu)}$$

35 where μ is the average number of times a TCR is observed in a sample. The average number of TCRs sequenced per sample in the Emerson data is 1.8×10^5 . Therefore the frequency of the TCR in the repertoire, in TCRs/million can be estimated by

$$Eq\ 2.\ f = -\log(p(X=0)) \times 10^6 / (1.8 \times 10^5)$$

40 As shown in Fig 3B the proportion of these SARS-COV-2 naïve individuals in whom we found each of the SARS-COV-2 expanded TCR is much greater than for non-expanded TCRs from the same repertoires. 2648 (52%) of the 5139 unique TCRbeta expanded sequences (the Emerson data is restricted to beta sequences) were observed two or more times in the Emerson data set. The estimated frequency distribution of these expanded TCRs estimated using equation 2 is plotted in Fig 3C, and compared to non-expanded TCRs. The mean frequency of the expanded set is 1.8 per million, over a thousand times higher than 1 per 10^9 , which as discussed above is the estimated frequency of individual TCR sequences in the naïve repertoire.

Fig 3

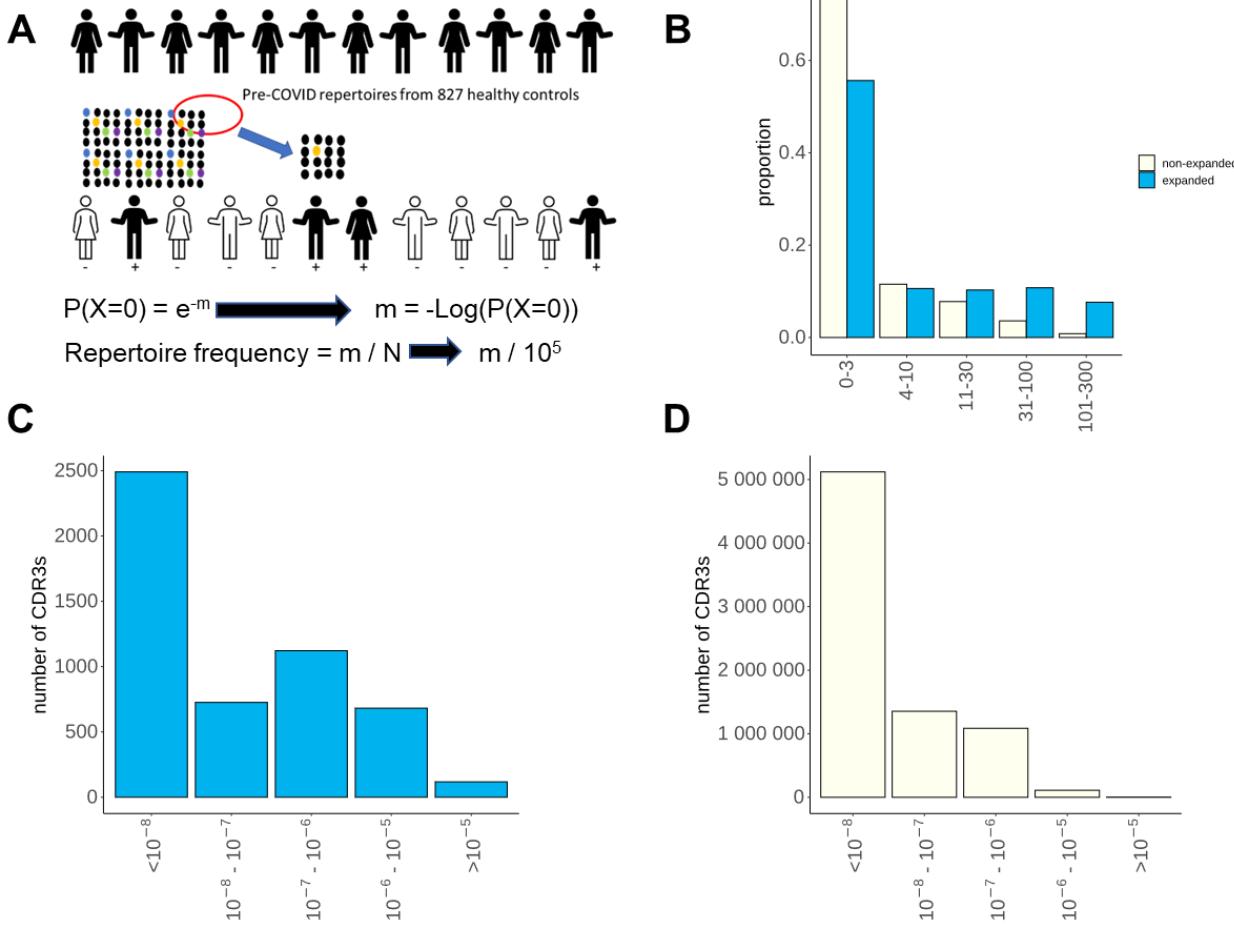


Fig 3. Sharing and frequency estimates of SARS-CoV-2 expanded TCR.

A Schematic of statistical inference of TCR frequencies from abundance data in pre-pandemic repertoires. m is the mean number of times a particular TCR is present in a given sample, given that that TCR is NOT detected in a proportion P of the 786 pre-pandemic repertoires examined. The estimated frequency is given by m/N , where N is the average number of TCRs sequenced in each sample.

B. Sharing of expanded TCRbeta or non-expanded control TCRs across the 786 repertoires of the Emerson data set, collected and sequenced several years before the SARS-CoV-2 pandemic. The x axis shows the number of individuals each TCR is observed in. The y axis shows the proportion of the expanded TCRs with a given sharing level.

C. The estimated frequency distribution of the 2648 expanded TCR beta sequences which are found in 2 or more individuals of the Emerson data set. The frequency of each expanded TCR was estimated using equation 2, as discussed in the text. TCR which were found less than twice in the Emerson data were assigned a frequency of $<10^{-6}$ and are represented by the column closest to the y axis.

D As for C, but for non-expanded TCRs from the SARS-CoV-2 repertoires.

Expanded TCRs in infected hosts cannot be explained by cross-reactivity to other coronaviruses

We have previously demonstrated that human TCR repertoire has a very broad range of TCR frequencies (41). High frequency precursor T cells could therefore arise either because they arise from naïve T cells with a higher than average frequency or due to cross-reaction (9, 42). The samples were collected during the first wave of SARS-CoV-2 in the UK, and the response could not therefore be attributed to pre-existing memory responses against SARS-CoV-2. We therefore explored the hypothesis that the

early wave of TCRs represented a restimulation of cross-reactive memory cells present as a result of exposure to common circulating human corona-viruses (hCoVs). In order to test this hypothesis, we analysed the 65 peptide epitopes which were recognized by the 141 TCRs which were functionally annotated as SARS-CoV-2 specific and were also identified as early expanders in our data set (as detailed in Fig 2). The similarity between these peptide epitopes and the homologous sequences from the major strains of circulating hCoVs is shown in Fig 4A. In general, with the exception of one NSP12 (polymerase) epitope, the SARS-CoV-2 sequences in the hCoVs showed substantial differences from their homologous sequences in SARS-CoV-2, making it unlikely that the TCRs specific for these sequences would recognize the equivalent peptide from the circulating hCoVs. In contrast, the epitopes recognized by these early annotated TCRs showed extremely high homology between the different variants of SARS-CoV-2 (Fig 4B). Within spike protein, for example, the early T cell epitopes were distinct from the mutations which are known to affect antibody recognition (Fig S13)(43). This analysis does not in any way preclude the existence of T cells which cross-react between SARS-CoV-2 and circulating hCoVs as suggested in other studies (9, 44). Rather, the analysis argues that the observed rapid expansion of SARS-CoV-2-specific TCRs can occur even in the absence of any such cross-reaction.

Fig 4

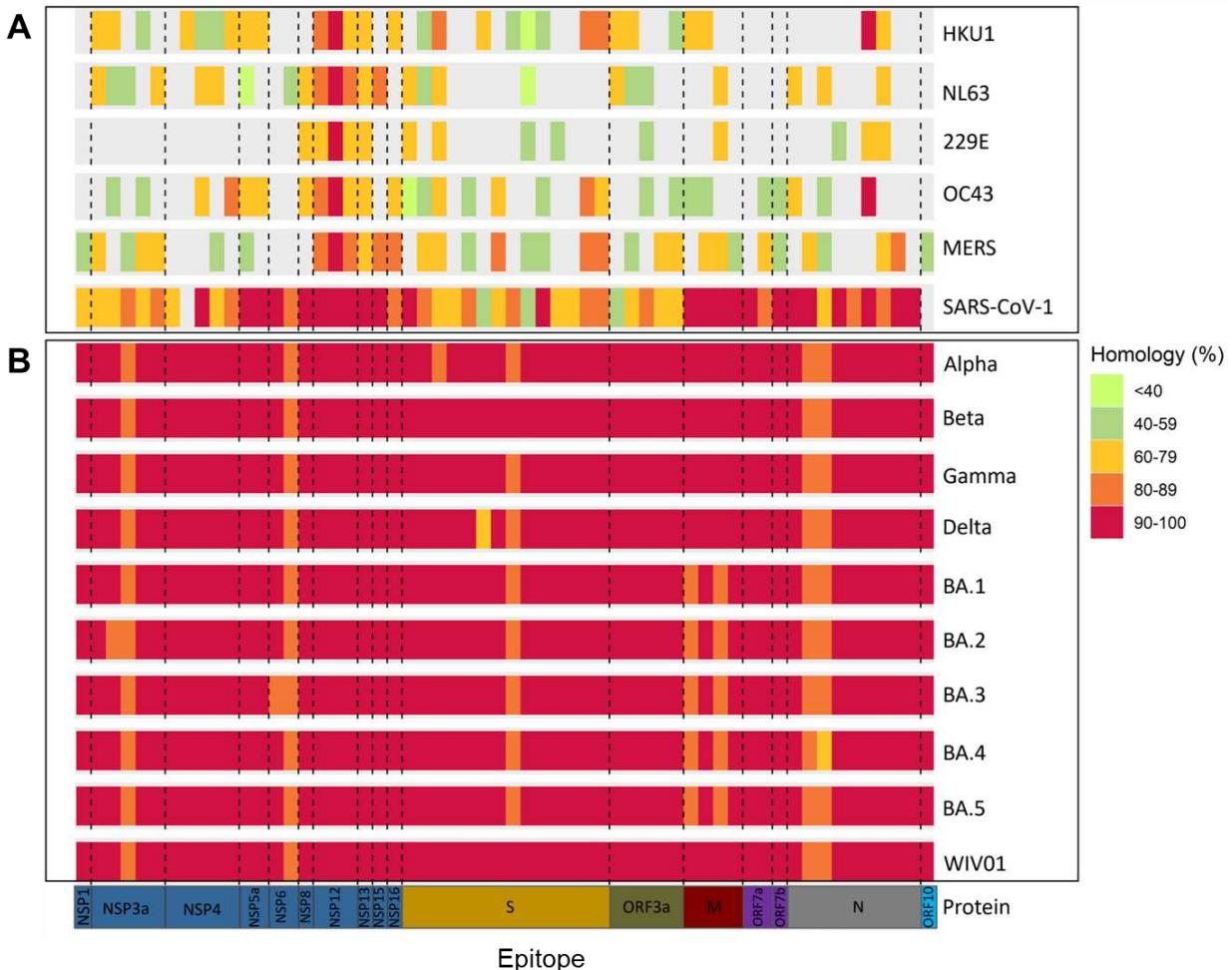


Fig 4. The homology between T cell epitopes of SARS-CoV-2 and circulating human coronaviruses A. Circulating coronaviruses
B Different strains of SARS-CoV-2

Each column is a functionally defined T cell epitope of SARS-CoV-2 which is recognised by a TCR whose sequence is found among the set of TCRs expanding following infection with the virus (as defined in Fig 1). The degree of homology is shown by colour coding.

5

High frequency LCMV specific TCRs in the naïve repertoire.

The number of viable cells collected at early timepoints of the COVIDSortium study limited the types of analysis we could carry out, and hence the ability to fully investigate the alternative hypothesis that the expanding TCRs we identified were derived from large pre-existing T cell clones in the naïve repertoire.

10 However, we reasoned that the phenomenon we observed for SARS-CoV-2 might be common to other viral infections, and might reflect fundamental features of the naïve T cell repertoire. We therefore

examined the abundances of TCRs from C57BL/6 mice exposed to LCMV. This model has been studied extensively, and several dominant and sub-dominant epitopes have been described (45–47). We

15 harvested spleens from mice at day 8 and day 40 post infection with saline or LCMV. In contrast to the SARS-CoV-2 data, we did not have to rely on dynamics to identify virus specific T cells. Instead we

directly isolated antigen specific T cells by sorting with fluorescent MHC tetramers bound to four well-described LCMV epitopes GP66, GP92, NP205 and NP396. An aliquot of spleen cells from each animal

was separately fractionated into naïve (CD62L+, CD44-), central memory (CD62+, CD44-) and effector (CD62-, CD44+) phenotype as described in Materials and Methods. We sequenced the TCR repertoire of

20 each antigen-specific and non-specific sample, and then searched for antigen specific TCR (tetramer-bound) sequences in the repertoire of the bulk naïve and effector populations (Fig5 A).

There was a strong enhancement in the abundance of epitope specific TCRs in the effector compartment of LCMV infected mice, especially at day 8 post infection, compared to saline injected mice (Fig 5B). In contrast, although many epitope specific TCRs were detected in the naïve repertoires, their abundance did not change between infected and PBS treated mice. Thus detection of epitope specific TCRs in the naïve repertoires is unlikely to arise from contamination during the fractionation step.

25 We then examined the frequency distribution of the epitope specific TCRs in the naïve repertoires (Fig 5C). To evaluate the difference between TCRs that expand early (which might have higher precursor frequency) and TCRs which expand later, we classified the TCRs as day 8 (already detected in the bulk

30 sequencing at day 8) or day 40 (not yet detected in bulk sequencing at day 8, but detected at day 40, Fig S14A). As shown in Fig 5C, in 3 out of 4 epitopes the early-rising antigen specific TCRs were enriched for high-abundance CDR3s, compared to the antigen-specific CDRs which arose at a later timepoint, or to a random sample of TCRs from the non-antigen specific repertoires (compare top four rows to bottom row).

35 Because we used a quantitative TCR sequencing protocol, we were able to estimate frequency by two orthogonal methods. In the left column, we show the frequencies estimated using the Poisson sampling approach which we developed for the SARS-CoV-2 data above. In brief, we determined the proportion of naïve repertoires in which we detected each TCR, and then used equation 2 to infer its frequency in the repertoire. Because we only had data from 10 mice, the granularity of this approach in this case was limited. In a second approach (right panels), we measured the average frequency of each

40 TCR directly (counts/total number of TCR in sample), and then averaged the frequencies over all 10 mice. TCRs not detected were assigned a frequency less than 10^{-6} in both methods. Both methods of

frequency estimation gave very similar results (Fig S14B), providing additional confidence in the statistical approach used for SARS-CoV-2. Overall, these results suggest that high frequency TCRs

specific for LCMV exist in the naïve repertoire prior to LCMV exposure.

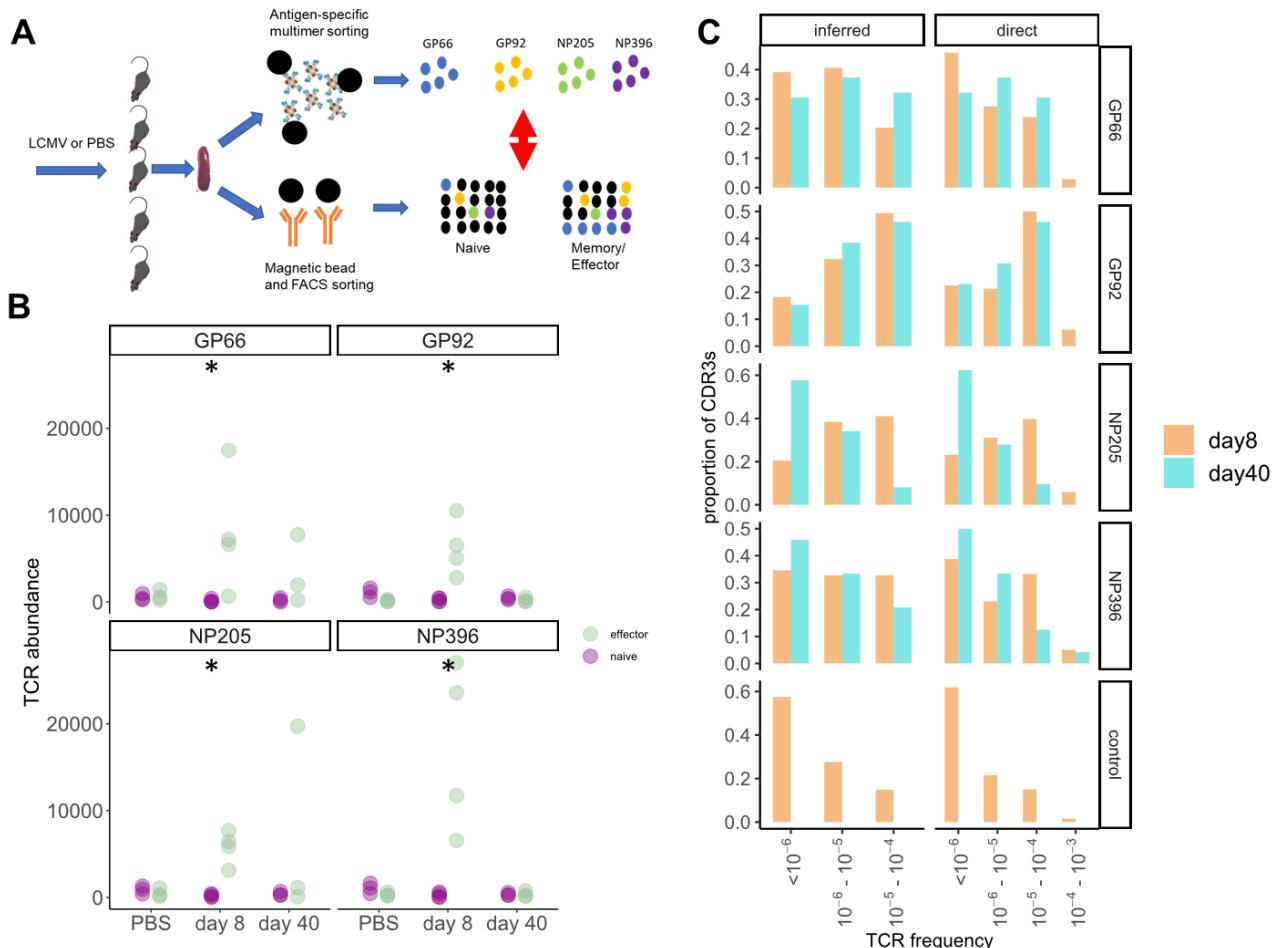


Fig 5. Abundance and frequency of LCMV specific TCR in naïve repertoires.

A schematic of the experimental design. Mice were injected with PBS or LCMV and spleen T cells harvested at day 8 and day 40 post-infection. Epitope-specific T cells were isolated using specific tetramer sorting, and different subpopulations of bulk T cells were fractionated by a combination of FACS sorting and magnetic bead/antibody sorting. Both sets of T cells were processed for TCR sequencing and the overlap between the populations analysed. B. Total abundance of epitope-specific TCRs in the naïve (purple) or effector (green) TCR repertoires in individual mice. The abundance of all four epitopes was larger in the effector population in the day 8 infected group, than in the PBS group (* $p < 0.001$, Mann Whitney). All other pairwise comparisons were not significant.

C Distribution histogram of frequency of epitope specific and control TCRs, estimated using either the statistical Poisson framework as in Fig 3 (left panels), or directly from TCR abundance data (right panels). Epitope-specific CDRs are split into CDRs that come up early (day 8, orange) and CDRs that come up later (day 40, cyan). The left column includes all epitope TCR which were not found within the bulk naïve population, and were assigned a frequency of less than 10^{-6} .

5

10

15

Discussion

The high prevalence and high degree of synchrony of the first wave of the SARS-COV-2 pandemic in the UK offered a rare opportunity to study the early dynamics of the T cell receptor repertoire following exposure to a natural novel viral infection. We identified a small number of TCRs which showed statistically significant changes in abundance within the first five weeks of our study, reasoning that viral exposure would lead to T cell clonal expansion and subsequent contraction. The set of TCRs identified in this way created a rapid transient wave of expansion, which peaked at or shortly after the first PCR+

SARS-COV-2 test. The number of expanded TCRs varied between individuals within a range of 18 to 912, and constituted only a very small fraction (1-2%) of the repertoire, consistent with the small proportion of activated T cells observed in blood using epitope specific tetramers or peptide specific T cell cytokine production (48, 49).

5 The identification of the responding TCRs was unbiased and based on dynamics alone. The wave of expansion was not observed in PCR- controls, consistent with a SARS-COV-2 specific response. However, the rapid expansion of TCRs might arise from cytokine-driven antigen non-specific as well as antigen specific clonal expansion. Access to a number of large sets of CDR3s annotated as SARS-COV-2 specific, as well as similar sets annotated for CMV and EBV specificity by functional assays allowed us to
10 interrogate the expanding set for antigen specificity. The expanded set of CDR3s were enriched for SARS-COV-2 specific annotated TCRs, although we cannot rule out that some bystander activation of other memory cells may occur. The annotated CDR3s observed in our data set shared V gene usage with the V gene of the corresponding annotated TCRs. Furthermore, most expanded annotated TCRs were found in individuals expressing the same HLA allele as the reported restriction element of the annotated
15 TCR. Both these observations build confidence in the functional annotation of the expanded TCRs as specific for SARS-COV-2.

20 The appearance of high frequency (> 100 TCRalpha/beta per million) T cells in the blood at the time of the first positive SARS-COV-2 PCR, or in some cases even before PCR+, needs to be considered in the context of the T cell precursor frequency prior to infection, and the known dynamics of T cell proliferation. The length of time between initial exposure and detectable SARS-COV-2 viremia (incubation period) has been estimated to be between 4 and 6 days (50–53). There is a lag between exposure and the arrival of antigen in the draining lymph node, and a further lag between TCR binding and the first cycle of T cell proliferation. Finally the time taken for a human T cell to complete the cell cycle is in the order of 12 - 14 hours (Phil Hodgkin, personal communication). There is therefore time for
25 no more than a maximum of 10 T cell divisions which, without taking into account cell death, allows each cell to proliferate 2^{10} (approximately 1,000-fold increase). The median frequency of the expanded TCRalpha and beta receptors prior to expansion must therefore be at least in the order of 1 in 10^7 . In a total pool of 10^{11} naïve T cells , this corresponds to a large (10,000) precursor T cell clone size.

30 In support of this hypothesis we were able to detect many of the expanded CDR3 sequences of the expanded TCRs in a significant proportion of pre-pandemic repertoires. We reasoned that TCRs which are present in many individuals (the public repertoire) must be present at high frequency in order to be observed in the small sample of blood which is typically sequenced. For example, in the Emerson data set of 786 repertoires collected and sequenced several years before the SARS-COV-2 pandemic, an average 100,000 TCR sequences are sequenced for each individual, out of an estimated total number of
35 $10^{11} - 10^{12}$ (54). We were therefore able to estimate the underlying frequency distribution of the 2648 expanded CDR3 TCRbeta sequences which we observed in two or more of the Emerson repertoires. As we predicted, the observed frequency distribution of these TCRs was between 10^{-7} and 10^{-5} , with an average frequency of $1.8 * 10^{-6}$. At least 50% of the TCRs which expanded therefore had high precursor frequency in pre-COVID repertoires of healthy individuals.

40 We considered two possible hypotheses to explain the presence of high precursor frequency SARS-COV-2 specific T cells in the pre-pandemic repertoire. The first hypothesis is that these TCRs are present on cross-reactive memory T cells specific for homologous epitopes on circulating human coronaviruses. Similar cross-reactive T cells, especially directed at the conserved polymerase, have recently been

described and may explain “abortive” infections with SARS-CoV-2(9). However, an analysis of the epitopes recognized by the annotated early expanding TCRs identified in this study (which are predominantly directed at structural proteins) showed little conservation between SARS-CoV-2 and the main circulating human coronaviruses (Fig 4). Therefore, while the existence of cross-reactive T cells is not in dispute, these cannot explain the early wave of SARS-CoV-2 specific expansion documented in this study. In contrast, the epitopes recognized by the early expanding T cells were highly conserved (in most cases identical) between all different strains of SARS-CoV-2. There is no evidence, therefore, that evolution has driven escape mutations within this set of T cell epitopes, similar to that seen for antibody. An important limitation of our study, however, is that the range of disease severity was very limited, and largely restricted to mild or asymptomatic disease. The potential role of early expanding T cells in limiting viral growth and hence pathology (55) cannot be investigated in this study, since all the infected HCW only developed mild disease .

The second hypothesis we explore is that the early SARS-CoV-2-specific T cell response we observe can be attributed to TCRs present at high precursor frequency in the naïve compartment. Recent estimates suggest that the naïve repertoire may comprise 10^8 - 10^9 different TCRs (56), but the frequency distribution is very broad with some TCRs present as high as 1 in 10^4 (41). Several factors may drive differential frequency of different T cell receptors in the circulation, although the mechanisms remain incompletely understood. We could not measure the frequency of individual TCRs in the pre-infection naïve repertoires of the HCW from this cohort since we did not have sufficient stored PBMC to sort and sequence the naïve compartment prior to infection.

We reasoned that the phenomenon of rapidly responding high frequency T cells in the naïve repertoire might be a more general feature of the T cell adaptive immune system. We therefore investigated one of the classical most well studied models of virus infection, LCMV (57). We isolated epitope specific T cells expanding very early after infection (day 8) and searched for TCRs with identical CDR3s in the naïve, as well as effector repertoires of both immunized and unimmunized mice. The results we obtained were very similar to those we had observed with SARS-CoV-2, but in this case we were able to directly test our principal hypothesis that T cells with high endogenous frequency are present within the naïve compartment, and contribute significantly to the early wave of T cell responses following infection. Moreover, we could detect a difference in the distribution of frequency in the naïve repertoire between cells that arise early and cells that are detected later in the infection, further supporting our hypothesis that early-expanding cells have higher precursor frequencies in the naïve compartment.

The design of the COVIDsortium study imposed some limitations. The protocol of the study, which was set up as rapidly as possible to capture the first wave of the SARS-CoV-2 pandemic in the UK, did not incorporate collection of sufficient number of viable cells at the early time points to carry out functional antigen-specific assays. Although we did not observe a wave of early TCR expansion/contraction in PCR-seronegative individuals, we cannot therefore exclude that some of the expanding TCRs were bystander memory cells to other antigens triggered by cytokine production. Furthermore, as discussed above, the same limitations on numbers of stored cells did not allow us to test directly our principal hypothesis that the early expanding TCRs arose from high frequency pre-exposure naïve T cell clones.

Despite these limitations, our study raises the intriguing possibility that the human immune system contains a population of high frequency precursor T cells which can provide a very rapid response to viral infections. These rapid responses do not require cross-reaction with circulating human coronaviruses, but are associated with pre-existing high frequency precursors. Further studies will be

required to determine whether these are high frequency naïve T cells, or represent random cross-reactivity of the memory pool. The ability to stimulate this rapid response may be an important factor in designing future vaccines to limit the growth and hence pathogenesis of respiratory viral infections.

5

References and Notes

1. M. D. T. Co, E. D. Kilpatrick, A. L. Rothman, Dynamics of the CD8 T-cell response following yellow fever virus 17D immunization. *Immunology*. **128** (2009), doi:10.1111/j.1365-2567.2009.03070.x.
2. C. L. Althaus, V. v. Ganusov, R. J. de Boer, Dynamics of CD8 + T Cell Responses during Acute and Chronic Lymphocytic Choriomeningitis Virus Infection . *The Journal of Immunology*. **179** (2007), doi:10.4049/jimmunol.179.5.2944.
10. 3. T. A. Treibel, C. Manisty, M. Burton, Á. McKnight, J. Lambourne, J. B. Augusto, X. Couto-Parada, T. Cutino-Moguel, M. Noursadeghi, J. C. Moon, COVID-19: PCR screening of asymptomatic health-care workers at London hospital. *The Lancet*. **395** (2020), , doi:10.1016/S0140-6736(20)31100-4.
15. 4. P. Moss, The T cell immune response against SARS-CoV-2. *Nat Immunol*. **23** (2022), , doi:10.1038/s41590-021-01122-w.
5. 5. A. Grifoni, J. Sidney, R. Vita, B. Peters, S. Crotty, D. Weiskopf, A. Sette, SARS-CoV-2 human T cell epitopes: Adaptive immune response against COVID-19. *Cell Host Microbe*. **29** (2021), , doi:10.1016/j.chom.2021.05.010.
20. 6. A. Tarke, J. Sidney, C. K. Kidd, J. M. Dan, S. I. Ramirez, E. D. Yu, J. Mateus, R. da Silva Antunes, E. Moore, P. Rubiro, N. Methot, E. Phillips, S. Mallal, A. Frazier, S. A. Rawlings, J. A. Greenbaum, B. Peters, D. M. Smith, S. Crotty, D. Weiskopf, A. Grifoni, A. Sette, Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep Med*. **2** (2021), doi:10.1016/j.xcrm.2021.100204.
25. 7. E. Stephenson, G. Reynolds, R. A. Botting, F. J. Calero-Nieto, M. D. Morgan, Z. K. Tuong, K. Bach, W. Sungnak, K. B. Worlock, M. Yoshida, N. Kumakawa, K. Kania, J. Engelbert, B. Olabi, J. S. Spegarova, N. K. Wilson, N. Mende, L. Jardine, L. C. S. Gardner, I. Goh, D. Horsfall, J. McGrath, S. Webb, M. W. Mather, R. G. H. Lindeboom, E. Dann, N. Huang, K. Polanski, E. Prigmore, F. Gothe, J. Scott, R. P. Payne, K. F. Baker, A. T. Hanrath, I. C. D. Schim van der Loeff, A. S. Barr, A. Sanchez-Gonzalez, L. Bergamaschi, F. Mescia, J. L. Barnes, E. Kilich, A. de Wilton, A. Saigal, A. Saleh, S. M. Janes, C. M. Smith, N. Gopee, C. Wilson, P. Coupland, J. M. Coxhead, V. Y. Kiselev, S. van Dongen, J. Bacardit, H. W. King, S. Baker, J. R. Bradley, G. Dougan, I. G. Goodfellow, R. K. Gupta, C. Hess, N. Kingston, P. J. Lehner, N. J. Matheson, W. H. Owehand, C. Saunders, K. G. C. Smith, C. Summers, J. E. D. Thaventhiran, M. Toshner, M. P. Weekes, A. Bucke, J. Calder, L. Canna, J. Domingo, A. Elmer, S. Fuller, J. Harris, S. Hewitt, J. Kennet, S. Jose, J. Kourampa, A. Meadows, C. O'Brien, J. Price, C. Publico, R. Rastall, C. Ribeiro, J. Rowlands, V. Ruffolo, H. Tordesillas, B. Bullman, B. J. Dunmore, S. Fawke, S. Gräf, J. Hodgson, C. Huang, K. Hunter, E. Jones, E. Legchenko, C. Matara, J. Martin, C. O'Donnell, L. Pointon, N. Pond, J. Shih, R. Sutcliffe, T. Tilly, C. Treacy, Z. Tong, J. Wood, M. Wylot, A. Betancourt, G. Bower, A. de Sa, M. Epping, O. Huhn, S. Jackson, I. Jarvis, J. Marsden, F. Nice, G. Okecha, O. Omarjee, M. Perera, N. Richoz, R. Sharma, L. Turner, E. M. D. D. de Bie, K. Bunclark, M. Josipovic, M. Mackay, A. Michael, S. Rossi, M. Selvan, S. Spencer, C. Yong, A. Ansaripour, L. Mwaura, C. Patterson, G. Polwarth, P. Polgarova, G. di Stefano, J. Allison, H. Butcher, D. Caputo,

5 D. Clapham-Riley, E. Dewhurst, A. Furlong, B. Graves, J. Gray, T. Ivers, M. Kasanicki, E. le Gresley, R. Linger, S. Meloy, F. Muldoon, N. Ovington, S. Papadia, I. Phelan, H. Stark, K. E. Stirrups, P. Townsend, N. Walker, J. Webster, A. J. Rostron, A. J. Simpson, S. Hambleton, E. Laurenti, P. A. Lyons, K. B. Meyer, M. Z. Nikolić, C. J. A. Duncan, K. G. C. Smith, S. A. Teichmann, M. R. Clatworthy, J. C. Marioni, B. Göttgens, M. Haniffa, Single-cell multi-omics analysis of the immune response in COVID-19. *Nat Med.* **27**, 904–916 (2021).

8. M. Yoshida, K. B. Worlock, N. Huang, R. G. H. Lindeboom, C. R. Butler, N. Kumasaka, C. Dominguez Conde, L. Mamanova, L. Bolt, L. Richardson, K. Polanski, E. Madissoon, J. L. Barnes, J. Allen-Hyttinen, E. Kilich, B. C. Jones, A. de Wilton, A. Wilbrey-Clark, W. Sungnak, J. P. Pett, J. Weller, E. Prigmore, H. Yung, P. Mehta, A. Saleh, A. Saigal, V. Chu, J. M. Cohen, C. Cane, A. Iordanidou, S. Shibuya, A. K. Reuschl, I. T. Herczeg, A. C. Argento, R. G. Wunderink, S. B. Smith, T. A. Poor, C. A. Gao, J. E. Dematte, G. R. S. Budinger, H. K. Donnelly, N. S. Markov, Z. Lu, G. Reynolds, M. Haniffa, G. S. Bowyer, M. Coates, M. R. Clatworthy, F. J. Calero-Nieto, B. Göttgens, C. O'Callaghan, N. J. Sebire, C. Jolly, P. de Coppi, C. M. Smith, A. v. Misharin, S. M. Janes, S. A. Teichmann, M. Z. Nikolić, K. B. Meyer, Local and systemic responses to SARS-CoV-2 infection in children and adults. *Nature*. **602**, 321–327 (2022).

10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85 90 95 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 896

9. L. Swadling, M. O. Diniz, N. M. Schmidt, O. E. Amin, A. Chandran, E. Shaw, C. Pade, J. M. Gibbons, N. le Bert, A. T. Tan, A. Jeffery-Smith, C. C. S. Tan, C. Y. L. Tham, S. Kucykowicz, G. Aidoo-Micah, J. Rosenheim, J. Davies, M. Johnson, M. P. Jensen, G. Joy, L. E. McCoy, A. M. Valdes, B. M. Chain, D. Goldblatt, D. M. Altmann, R. J. Boyton, C. Manisty, T. A. Treibel, J. C. Moon, H. Abbass, A. Abiodun, M. Alfarih, Z. Alldis, M. Andiapan, J. Artico, J. B. Augusto, G. L. Baca, S. N. L. Bailey, A. N. Bhuva, A. Boulter, R. Bowles, R. J. Boyton, O. v. Bracken, B. O'Brien, T. Brooks, N. Bullock, D. K. Butler, G. Captur, N. Champion, C. Chan, D. Collier, J. C. de Sousa, X. Couto-Parada, T. Cutino-Moguel, R. H. Davies, B. Douglas, C. di Genova, K. Dieobi-Anene, A. Ellis, K. Feehan, M. Finlay, M. Fontana, N. Forooghi, C. Gaier, D. Gilroy, M. Hamblin, G. Harker, J. Hewson, L. M. Hickling, A. D. Hingorani, L. Howes, A. Hughes, G. Hughes, R. Hughes, I. Itua, V. Jardim, W. Y. J. Lee, M. petra Jensen, J. Jones, M. Jones, G. Joy, V. Kapil, H. Kurdi, J. Lambourne, K. M. Lin, S. Louth, V. Mandadapu, Á. McKnight, K. Menacho, C. Mfuko, O. Mitchelmore, C. Moon, S. M. Murray, M. Noursadeghi, A. Otter, S. Palma, R. Parker, K. Patel, B. Pawarova, S. E. Petersen, B. Piniera, F. P. Pieper, D. Pope, M. Prossora, L. Rannigan, A. Rapala, C. J. Reynolds, A. Richards, M. Robathan, G. Sambile, A. Semper, A. Seraphim, M. Simion, A. Smit, M. Sugimoto, S. Taylor, N. Temperton, S. Thomas, G. D. Thornton, A. Tucker, J. Veerapen, M. Vijayakumar, S. Welch, T. Wodehouse, L. Wynne, D. Zahedi, L. van Dorp, F. Balloux, Á. McKnight, M. Noursadeghi, A. Bertoletti, M. K. Maini, Pre-existing polymerase-specific T cells expand in abortive seronegative SARS-CoV-2. *Nature* (2021), doi:10.1038/s41586-021-04186-8.

10. K. Kedzierska, P. G. Thomas, Count on us: T cells in SARS-CoV-2 infection and vaccination. *Cell Rep Med.* **3** (2022), doi:10.1016/J.XCRM.2022.100562.

11. D. Goldblatt, G. Alter, S. Crotty, S. A. Plotkin, Correlates of protection against SARS-CoV-2 infection and COVID-19 disease. *Immunol Rev* (2022), doi:10.1111/IMR.13091.

12. J. S. Heitmann, T. Bilich, C. Tandler, A. Nelde, Y. Maringer, M. Marconato, J. Reusch, S. Jäger, M. Denk, M. Richter, L. Anton, L. M. Weber, M. Roerden, J. Bauer, J. Rieth, M. Wacker, S. Hörber, A. Peter, C. Meisner, I. Fischer, M. W. Löffler, J. Karbach, E. Jäger, R. Klein, H. G. Rammensee, H. R.

Salih, J. S. Walz, A COVID-19 peptide vaccine for the induction of SARS-CoV-2 T cell immunity. *Nature*. **601** (2022), doi:10.1038/s41586-021-04232-5.

13. A. A. Minervina, E. A. Komech, A. Titov, M. B. Koraichi, E. Rosati, I. Z. Mamedov, A. Franke, G. A. Efimov, D. M. Chudakov, T. Mora, A. M. Walczak, Y. B. Lebedev, M. v. Pogorelyy, Longitudinal high-throughput tcr repertoire profiling reveals the dynamics of t-cell memory formation after mild covid-19 infection. *Elife*. **10** (2021), doi:10.7554/elife.63502.

5 14. T. M. Wilkinson, C. K. F. Li, C. S. C. Chui, A. K. Y. Huang, M. Perkins, J. C. Liebner, R. Lambkin-Williams, A. Gilbert, J. Oxford, B. Nicholas, K. J. Staples, T. Dong, D. C. Douek, A. J. McMichael, X. N. Xu, Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans. *Nat Med*. **18**, 274–280 (2012).

10 15. A. Chandran, J. Rosenheim, G. Nageswaran, L. Swadling, G. Pollara, R. K. Gupta, A. R. Burton, J. A. Guerra-Assunção, A. Woolston, T. Ronel, C. Pade, J. M. Gibbons, B. Sanz-Magallon Duque De Estrada, M. Robert de Massy, M. Whelan, A. Semper, T. Brooks, D. M. Altmann, R. J. Boyton, Á. McKnight, G. Captur, C. Manisty, T. A. Treibel, J. C. Moon, G. S. Tomlinson, M. K. Maini, B. M. Chain, M. Noursadeghi, Rapid synchronous type 1 IFN and virus-specific T cell responses characterize first wave non-severe SARS-CoV-2 infections. *Cell Rep Med*. **3** (2022), doi:10.1016/j.xcrm.2022.100557.

15 16. A. Chandran, J. Rosenheim, G. Nageswaran, L. Swadling, G. Pollara, R. K. Gupta, A. R. Burton, J. A. Guerra-Assunção, A. Woolston, T. Ronel, C. Pade, J. M. Gibbons, B. Sanz-Magallon Duque De Estrada, M. Robert de Massy, M. Whelan, A. Semper, T. Brooks, D. M. Altmann, R. J. Boyton, Á. McKnight, G. Captur, C. Manisty, T. A. Treibel, J. C. Moon, G. S. Tomlinson, M. K. Maini, B. M. Chain, M. Noursadeghi, Rapid synchronous type 1 IFN and virus-specific T cell responses characterize first wave non-severe SARS-CoV-2 infections. *Cell Rep Med*. **3** (2022), doi:10.1016/j.xcrm.2022.100557.

20 17. C. J. Reynolds, L. Swadling, J. M. Gibbons, C. Pade, M. P. Jensen, M. O. Diniz, N. M. Schmidt, D. K. Butler, O. E. Amin, S. N. L. Bailey, S. M. Murray, F. P. Pieper, S. Taylor, J. Jones, M. Jones, W. Y. J. Lee, J. Rosenheim, A. Chandran, G. Joy, C. di Genova, N. Temperton, J. Lambourne, T. Cutino-Moguel, M. Andiapen, M. Fontana, A. Smit, A. Semper, B. O'Brien, B. Chain, T. Brooks, C. Manisty, T. Treibel, J. C. Moon, M. Noursadeghi, D. M. Altmann, M. K. Maini, Á. McKnight, R. J. Boyton, H. Abbass, A. Abiodun, M. Alfarih, Z. Alldis, D. M. Altmann, O. E. Amin, M. Andiapen, J. Artico, J. B. Augusto, G. L. Baca, S. N. L. Bailey, A. N. Bhuvan, A. Boulter, R. Bowles, R. J. Boyton, O. v. Bracken, B. O'Brien, T. Brooks, N. Bullock, D. K. Butler, G. Captur, N. Champion, C. Chan, A. Chandran, J. C. de Sousa, X. Couto-Parada, T. Cutino-Moguel, R. H. Davies, B. Douglas, C. di Genova, K. Dieobi-Anene, M. O. Diniz, K. Feehan, M. Finlay, M. Fontana, N. Forooghi, J. M. Gibbons, D. Gilroy, M. Hamblin, J. Hewson, L. M. Hickling, A. D. Hingorani, L. Howes, I. Itua, V. Jardim, M. Jensen, J. Jones, M. Jones, G. Joy, V. Kapil, H. Kurdi, J. Lambourne, S. Louth, M. K. Maini, V. Mandadapu, C. Manisty, Á. McKnight, K. Menacho, C. Mfuko, O. Mitchelmore, C. Moon, J. Moon, S. M. Murray, M. Noursadeghi, A. Otter, C. Pade, S. Palma, R. Parker, K. Patel, M. Pawarova, S. E. Petersen, B. Piniera, F. P. Pieper, L. Rannigan, A. Rapala, C. J. Reynolds, A. Richards, M. Robathan, J. Rosenheim, G. Sambile, N. M. Schmidt, A. Semper, A. Seraphim, M. Simion, A. Smit, M. Sugimoto, L. Swadling, S. Taylor, N. Temperton, S. Thomas, G. D. Thornton, T. Treibel, A. Tucker, J. Veerapen, M. Vijayakumar, T. Warner, S. Welch, T. Wodehouse, L. Wynne, D. Zahedi, Discordant

neutralizing antibody and T cell responses in asymptomatic and mild SARS-CoV-2 infection. *Sci Immunol.* **5** (2020), doi:10.1126/SCIIMMUNOL.ABF3698.

18. S. Astbury, C. J. Reynolds, D. K. Butler, D. C. Muñoz-Sandoval, K. M. Lin, F. P. Pieper, A. Otter, A. Kouraki, L. Cusin, J. Nightingale, A. Vijay, S. Craxford, G. P. Aithal, P. J. Tighe, J. M. Gibbons, C. Pade, G. Joy, M. Maini, B. Chain, A. Semper, T. Brooks, B. J. Ollivere, Á. McKnight, M. Noursadeghi, T. A. Treibel, C. Manisty, J. C. Moon, A. M. Valdes, R. J. Boyton, D. M. Altmann, H. Abbass, A. Abiodun, M. Alfarah, Z. Alldis, D. M. Altmann, O. E. Amin, M. Andiapan, J. Artico, J. B. Augusto, G. L. Baca, S. N. L. Bailey, A. N. Bhuva, A. Boulter, R. Bowles, R. J. Boyton, O. v. Bracken, B. O'Brien, T. Brooks, N. Bullock, D. K. Butler, G. Captur, N. Champion, C. Chan, A. Chandran, D. Collier, J. C. de Sousa, X. Couto-Parada, T. Cutino-Moguel, R. H. Davies, B. Douglas, C. Genova, K. Dieobi-Anene, M. O. Diniz, A. Ellis, K. Feehan, M. Finlay, M. Fontana, N. Forooghi, C. Gaier, J. M. Gibbons, D. Gilroy, M. Hamblin, G. Harker, J. Hewson, L. M. Hickling, A. D. Hingorani, L. Howes, A. Hughes, G. Hughes, R. Hughes, I. Itua, V. Jardim, W. Y. J. Lee, M. Jensen, J. Jones, M. Jones, G. Joy, V. Kapil, H. Kurdi, J. Lambourne, K. M. Lin, S. Louth, M. K. Maini, V. Mandadapu, C. Manisty, Á. McKnight, K. Menacho, C. Mfuko, O. Mitchelmore, C. Moon, J. C. Moon, D. C. Muñoz-Sandoval, S. M. Murray, M. Noursadeghi, A. Otter, C. Pade, S. Palma, R. Parker, K. Patel, B. Pawarova, S. E. Petersen, B. Piniera, F. P. Pieper, D. Pope, M. Prossora, L. Rannigan, A. Rapala, C. J. Reynolds, A. Richards, M. Robathan, J. Rosenheim, G. Sambile, N. M. Schmidt, A. Semper, A. Seraphim, M. Simion, A. Smit, M. Sugimoto, L. Swadling, S. Taylor, N. Temperton, S. Thomas, G. D. Thornton, T. A. Treibel, A. Tucker, J. Veerapen, M. Vijayakumar, S. Welch, T. Wodehouse, L. Wynne, D. Zahedi, D. M. Altmann, R. J. Boyton, T. Brooks, B. Chain, M. K. Maini, C. Manisty, Á. McKnight, J. C. Moon, M. Noursadeghi, T. A. Treibel, G. P. Aithal, W. Ashraf, S. Astbury, J. K. Ball, J. G. Chappell, S. Craxford, L. M. L. Cusin, J. D. Duncan, A. Ikram, W. L. Irving, H. J. Jackson, A. Kelly, M. Lingaya, B. A. Marson, J. Newham, J. Nightingale, A. Norrish, B. Nowicka, B. J. Ollivere, A. W. Tarr, P. J. Tighe, T. Tsoleridis, R. A. Urbanowicz, A. M. Valdes, A. Vijay, HLA-DR polymorphism in SARS-CoV-2 infection and susceptibility to symptomatic COVID-19. *Immunology.* **166** (2022), doi:10.1111/IMM.13450.

19. I. Uddin, K. Joshi, T. Oakes, J. M. Heather, C. Swanton, B. Chain, *An economical, quantitative, and robust protocol for high-throughput T cell receptor sequencing from tumor or blood* (2019), vol. 1884.

20. T. Oakes, J. M. Heather, K. Best, R. Byng-Maddick, C. Husovsky, M. Ismail, K. Joshi, G. Maxwell, M. Noursadeghi, N. Riddell, T. Ruehl, C. T. Turner, I. Uddin, B. Chain, Quantitative Characterization of the T Cell Receptor Repertoire of Naïve and Memory Subsets Using an Integrated Experimental and Computational Pipeline Which Is Robust, Economical, and Versatile. *Front Immunol.* **8** (2017), doi:10.3389/FIMMU.2017.01267.

21. A. Karatzoglou, A. Smola, K. Hornik, A. Zeileis, **kernlab** - An S4 Package for Kernel Methods in *R. J Stat Softw.* **11**, 1–20 (2004).

22. G. Csardi, T. Nepusz, The igraph software package for complex network research. *InterJournal. Complex Systems*, 1695 (2006).

23. Y. Peng, S. L. Felce, D. Dong, F. Penkava, A. J. Mentzer, X. Yao, G. Liu, Z. Yin, J. L. Chen, Y. Lu, D. Wellington, P. A. C. Wing, D. C. C. Dominey-Foy, C. Jin, W. Wang, M. A. Hamid, R. A. Fernandes, B. Wang, A. Fries, X. Zhuang, N. Ashley, T. Rostron, C. Waugh, P. Sopp, P. Hublitz, R. Beveridge, T. K.

Tan, C. Dold, A. J. Kwok, C. Rich-Griffin, W. Dejnirattisa, C. Liu, P. Kurupati, I. Nassiri, R. A. Watson, O. Tong, C. A. Taylor, P. Kumar Sharma, B. Sun, F. Curion, S. Revale, L. C. Garner, K. Jansen, R. C. Ferreira, M. Attar, J. W. Fry, R. A. Russell, H. J. Stauss, W. James, A. Townsend, L. P. Ho, P.

5 Klenerman, J. Mongkolsapaya, G. R. Scream, C. Dendrou, S. N. Sansom, R. Bashford-Rogers, B. Chain, G. L. Smith, J. A. McKeating, B. P. Fairfax, P. Bowness, A. J. McMichael, G. Ogg, J. C. Knight, T. Dong, An immunodominant NP105–113-B*07:02 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease. *Nature Immunology* 2021 23:1. 23, 50–61 (2021).

24. D. A. Bolotin, S. Poslavsky, I. Mitrophanov, M. Shugay, I. Z. Mamedov, E. v Putintseva, D. M. Chudakov, MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 12, 380–381 (2015).

10 25. Y. C. Peng, B. Wang, K. Talaat, R. Karron, T. J. Powell, H. Zeng, D. Dong, C. J. Luke, A. McMichael, K. Subbarao, T. Dong, Boosted influenza-specific T cell responses after H5N1 pandemic live attenuated influenza virus vaccination. *Front Immunol*. 6, 287 (2015).

15 26. C. C. S. Tan, C. J. Owen, C. Y. L. Tham, A. Bertoletti, L. van Dorp, F. Balloux, Pre-existing T cell-mediated cross-reactivity to SARS-CoV-2 cannot solely be explained by prior exposure to endemic human coronaviruses. *Infection, Genetics and Evolution*. 95, 105075 (2021).

27. S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall*. 1, 33–46 (2017).

20 28. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*. 22, 30494 (2017).

29. K. Katoh, K. Misawa, K. I. Kuma, T. Miyata, MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 30, 3059–3066 (2002).

30 30. J. Huddleston, J. Hadfield, T. Sibley, J. Lee, K. Fay, M. Ilcisin, E. Harkins, T. Bedford, R. Neher, E. Hodcroft, Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw*. 6, 2906 (2021).

31. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*. 20, 289–290 (2004).

32. R. O. Emerson, W. S. DeWitt, M. Vignali, J. Gravley, J. K. Hu, E. J. Osborne, C. Desmarais, M. Klinger, C. S. Carlson, J. A. Hansen, M. Rieder, H. S. Robins, Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet*. 49, 659–665 (2017).

33. M. Mark, S. Reich-Zeliger, E. Greenstein, D. Reshef, A. Madi, B. Chain, N. Friedman, Antigen experience relaxes the organisational structure of the T cell receptor repertoire, doi:10.1101/2021.10.26.465870.

35 34. C. Manisty, T. A. Treibel, M. Jensen, A. Semper, G. Joy, R. K. Gupta, T. Cutino-Moguel, M. Andiapen, J. Jones, S. Taylor, A. Otter, C. Pade, J. Gibbons, J. Lee, J. Bacon, S. Thomas, C. Moon, M. Jones, D. Williams, J. Lambourne, M. Fontana, D. M. Altmann, R. Boyton, M. Maini, A. McKnight, B. Chain, M. Noursadeghi, J. C. Moon, Time series analysis and mechanistic modelling

of heterogeneity and sero-reversion in antibody responses to mild SARS-CoV-2 infection. *EBioMedicine*. **65** (2021), doi:10.1016/j.ebiom.2021.103259.

35. M. Shugay, D. v. Bagaev, I. v. Zvyagin, R. M. Vroomans, J. C. Crawford, G. Dolton, E. A. Komech, A. L. Sycheva, A. E. Koneva, E. S. Egorov, A. v. Eliseev, E. van Dyk, P. Dash, M. Attaf, C. Rius, K. Ladell, J. E. McLaren, K. K. Matthews, E. B. Clemens, D. C. Douek, F. Luciani, D. van Baarle, K. Kedzierska, C. Kesmir, P. G. Thomas, D. A. Price, A. K. Sewell, D. M. Chudakov, VDJdb: A curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* **46** (2018), doi:10.1093/nar/gkx760.

5 36. J. M. Francis, D. Leistritz-Edwards, A. Dunn, C. Tarr, J. Lehman, C. Dempsey, A. Hamel, V. Rayon, G. Liu, Y. Wang, M. Wille, M. Durkin, K. Hadley, A. Sheena, B. Roscoe, M. Ng, G. Rockwell, M. Manto, E. Gienger, J. Nickerson, MGH COVID-19 Collection and Processing Team2‡, A. Moarefi, M. Noble, T. Malia, P. D. Bardwell, W. Gordon, J. Swain, M. Skoberne, K. Sauer, T. Harris, A. W. Goldrath, A. K. Shalek, A. J. Coyle, C. Benoist, D. C. Pregibon, N. Jilg, J. Li, A. Rosenthal, C. Wong, G. Daley, D. Golan, H. Heller, A. Sharpe, B. A. Abayneh, P. Allen, D. Antille, K. Armstrong, S. Boyce, J. Braley, K. Branch, K. Broderick, J. Carney, A. Chan, S. Davidson, M. Dougan, D. Drew, A. Elliman, K. Flaherty, J. Flannery, P. Forde, E. Gettings, A. Griffin, S. Grimmel, K. Grinke, K. Hall, M. Healy, D. Henault, G. Holland, C. Kayitesi, V. LaValle, Y. Lu, S. Luthern, J. M. Schneider, B. Martino, R. McNamara, C. Nambu, S. Nelson, M. Noone, C. Ommerborn, L. C. Pacheco, N. Phan, F. A. Porto, E. Ryan, K. Selleck, S. Slaughenhaupt, K. S. Sheppard, E. Suschana, V. Wilson, M. Carrington, M. Martin, Y. Yuki, G. Alter, A. Balazs, J. Bals, M. Barbash, Y. Bartsch, J. Boucau, M. Carrington, J. Chevalier, F. Chowdhury, E. DeMers, K. Einkauf, J. Fallon, L. Fedirko, K. Finn, P. Garcia-Broncano, M. S. Ghebremichael, C. Hartana, C. Jiang, K. Judge, P. Kaplonek, M. Karpell, P. Lai, E. C. Lam, K. Lefteri, X. Lian, M. Lichtenfeld, D. Lingwood, H. Liu, J. Liu, N. Ly, Z. M. Hill, A. Michell, I. Millstrom, N. Miranda, C. O'Callaghan, M. Osborn, S. Pillai, Y. Rassadkina, A. Reissis, F. Ruzicka, K. Seiger, L. Sessa, C. Sharr, S. Shin, N. Singh, W. Sun, X. Sun, H. Ticheli, A. Trocha-Piechocka, B. Walker, D. Worrall, X. G. Yu, A. Zhu, Allelic variation in class I HLA determines CD8 + T cell repertoire shape and cross-reactive memory responses to SARS-CoV-2. *Sci Immunol.* **7**, eabk3070 (2022).

10 37. P. Dash, A. J. Fiore-Gartland, T. Hertz, G. C. Wang, S. Sharma, A. Souquette, J. C. Crawford, E. B. Clemens, T. H. O. Nguyen, K. Kedzierska, N. L. la Gruta, P. Bradley, P. G. Thomas, Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature*. **547**, 89–93 (2017).

15 38. J. Glanville, H. Huang, A. Nau, O. Hatton, L. E. Wagar, F. Rubelt, X. Ji, A. Han, S. M. Krams, C. Pettus, N. Haas, C. S. L. Arlehamn, A. Sette, S. D. Boyd, T. J. Scriba, O. M. Martinez, M. M. Davis, Identifying specificity groups in the T cell receptor repertoire. *Nature*. **547**, 94–98 (2017).

20 39. K. Mayer-Blackwell, S. Schattgen, L. Cohen-Lavi, J. C. Crawford, A. Souquette, J. A. Gaever, T. Hertz, P. G. Thomas, P. G. Bradley, A. Fiore-Gartland, Tcr meta-clonotypes for biomarker discovery with tcrdist3 enabled identification of public, hla-restricted clusters of sars-cov-2 tcrs. *Elife*. **10** (2021), doi:10.7554/elife.68605.

25 40. K. Joshi, M. Robert de Massy, M. Ismail, J. L. Reading, I. Uddin, A. Woolston, E. Hatipoglu, T. Oakes, R. Rosenthal, T. Peacock, T. Ronel, M. Noursadeghi, V. Turati, A. J. S. Furness, A. Georgiou, Y. N. S. Wong, A. ben Aissa, M. Werner Sunderland, M. Jamal-Hanjani, S. Veeriah, N. J. Birkbak, G. A. Wilson, C. T. Hiley, E. Ghorani, J. A. Guerra-Assunção, J. Herrero, T. Enver, S. R. Hadrup, A. Hackshaw, K. S. Peggs, N. McGranahan, C. Swanton, TRACERx Consortium, S. A. Quezada, B.

Chain, Spatial heterogeneity of the T cell receptor repertoire reflects the mutational landscape in lung cancer. *Nat Med.* **25**, 1549–1559 (2019).

41. P. C. de Greef, T. Oakes, B. Gerritsen, M. Ismail, J. M. Heather, R. Hermsen, B. Chain, R. J. de Boer, The naive t-cell receptor repertoire has an extremely broad distribution of clone sizes. *eLife.* **9** (2020), doi:10.7554/eLife.49900.

5

42. L. Bartolo, S. Afroz, Y.-G. Pan, R. Xu, L. Williams, C.-F. Lin, C. Tanes, K. Bittinger, E. S. Friedman, P. A. Gimotty, G. D. Wu, L. F. Su, SARS-CoV-2-specific T cells in unexposed adults display broad trafficking potential and cross-react with commensal antigens. *Sci Immunol* (2022), doi:10.1126/SCIMMUNOL.ABN3127.

10

43. D. M. Altmann, C. J. Reynolds, R. J. Boyton, SARS-CoV-2 variants: Subversion of antibody response and predicted impact on T cell recognition. *Cell Rep Med.* **2** (2021), doi:10.1016/j.xcrm.2021.100286.

15

44. A. Becerra-Artiles, J. M. Calvo-Calle, M. Co, P. Nanaware, J. Cruz, G. C. Weaver, L. Lu, C. Forcini, R. W. Finberg, A. Moormann, L. J. Stern, Broadly-recognized, cross-reactive SARS-CoV-2 CD4 T cell epitopes are highly conserved across human coronaviruses and presented by common HLA alleles. *bioRxiv* (2022).

20

45. R. G. van der Most, K. Murali-Krishna, J. L. Whitton, C. Oseroff, J. Alexander, S. Southwood, J. Sidney, R. W. Chesnut, A. Sette, R. Ahmed, Identification of Db- and Kb-restricted subdominant cytotoxic T-cell responses in lymphocytic choriomeningitis virus-infected mice. *Virology.* **240**, 158–167 (1998).

25

46. J. E. Gairin, H. Mazarguil, D. Hudrisier, M. B. Oldstone, Optimal lymphocytic choriomeningitis virus sequences restricted by H-2Db major histocompatibility complex class I molecules and presented to cytotoxic T lymphocytes. *J Virol.* **69**, 2297–2305 (1995).

47. C. Dow, C. Oseroff, B. Peters, C. Nance-Sotelo, J. Sidney, M. Buchmeier, A. Sette, B. R. Mothé, Lymphocytic choriomeningitis virus infection yields overlapping CD4+ and CD8+ T-cell responses. *J Virol.* **82**, 11734–11741 (2008).

30

48. A. A. Minervina, M. v. Pogorelyy, A. M. Kirk, J. C. Crawford, E. K. Allen, C.-H. Chou, R. C. Mettelman, K. J. Allison, C.-Y. Lin, D. C. Brice, X. Zhu, K. Vegesana, G. Wu, S. Trivedi, P. Kottapalli, D. Darnell, S. McNeely, S. R. Olsen, S. Schultz-Cherry, J. H. Estepp, A. Gaur, J. Hoffman, M. Mori, L. Tang, E. Tuomanen, R. Webby, H. Hakim, R. T. Hayden, D. R. Hijano, R. Bajracharya, W. Awad, L.-A. van de Velde, B. L. Clark, T. L. Wilson, A. Souquette, A. Castellaw, R. H. Dallas, J. Hodges, A. Gowen, J. Russell-Bell, J. Sparks, D. E. Wittman, T. P. Fabrizio, S. Cherry, E. K. Roubidoux, V. Cortez, P. Freiden, N. Wohlgemuth, K. Whitt, M. A. McGargill, J. Wolf, P. G. Thomas, SARS-CoV-2 antigen exposure history shapes phenotypes and specificity of memory CD8+ T cells. *Nat Immunol.* **23**, 781–790 (2022).

35

49. C. J. Reynolds, L. Swadling, J. M. Gibbons, C. Pade, M. P. Jensen, M. O. Diniz, N. M. Schmidt, D. K. Butler, O. E. Amin, S. N. L. Bailey, S. M. Murray, F. P. Pieper, S. Taylor, J. Jones, M. Jones, W. Y. J. Lee, J. Rosenheim, A. Chandran, G. Joy, C. di Genova, N. Temperton, J. Lambourne, T. Cutino-Moguel, M. Andiapen, M. Fontana, A. Smit, A. Semper, B. O'Brien, B. Chain, T. Brooks, C. Manisty, T. Treibel, J. C. Moon, M. Noursadeghi, D. M. Altmann, M. K. Maini, Á. McKnight, R. J. Boyton, H.

40

Abbass, A. Abiodun, M. Alfarih, Z. Alldis, D. M. Altmann, O. E. Amin, M. Andiapen, J. Artico, J. B. Augusto, G. L. Baca, S. N. L. Bailey, A. N. Bhuva, A. Boulter, R. Bowles, R. J. Boyton, O. v. Bracken, B. O'Brien, T. Brooks, N. Bullock, D. K. Butler, G. Captur, N. Champion, C. Chan, A. Chandran, J. C. de Sousa, X. Couto-Parada, T. Cutino-Moguel, R. H. Davies, B. Douglas, C. di Genova, K. Dieobi-Anene, M. O. Diniz, K. Feehan, M. Finlay, M. Fontana, N. Forooghi, J. M. Gibbons, D. Gilroy, M. Hamblin, J. Hewson, L. M. Hickling, A. D. Hingorani, L. Howes, I. Itua, V. Jardim, M. Jensen, J. Jones, M. Jones, G. Joy, V. Kapil, H. Kurdi, J. Lambourne, S. Louth, M. K. Maini, V. Mandadapu, C. Manisty, Á. McKnight, K. Menacho, C. Mfuko, O. Mitchelmore, C. Moon, J. Moon, S. M. Murray, M. Noursadeghi, A. Otter, C. Pade, S. Palma, R. Parker, K. Patel, M. Pawarova, S. E. Petersen, B. Piniera, F. P. Pieper, L. Rannigan, A. Rapala, C. J. Reynolds, A. Richards, M. Robathan, J. Rosenheim, G. Sambile, N. M. Schmidt, A. Semper, A. Seraphim, M. Simion, A. Smit, M. Sugimoto, L. Swadling, S. Taylor, N. Temperton, S. Thomas, G. D. Thornton, T. Treibel, A. Tucker, J. Veerapen, M. Vijayakumar, T. Warner, S. Welch, T. Wodehouse, L. Wynne, D. Zahedi, Discordant neutralizing antibody and T cell responses in asymptomatic and mild SARS-CoV-2 infection. *Sci Immunol.* **5** (2020), doi:10.1126/SCIIMMUNOL.ABF3698.

50. J. Bernal Lopez, N. Panagiotopoulos, C. Byers, T. G. Vilaplana, N. Boddington, X. S. Zhang, A. Charlett, S. Elgohari, L. Coughlan, R. Whillock, S. Logan, H. Bolt, M. Sinnathamby, L. Letley, P. MacDonald, R. Vivancos, O. Edeghere, C. Anderson, K. Paranthaman, S. Cottrell, J. McMenamin, M. Zambon, G. Dabrera, M. Ramsay, V. Saliba, Transmission dynamics of COVID-19 in household and community settings in the United Kingdom, January to March 2020. *Eurosurveillance.* **27**, 1 (2022).

51. S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, J. Lessler, The incubation period of coronavirus disease 2019 (CoVID-19) from publicly reported confirmed cases: Estimation and application. *Ann Intern Med.* **172**, 577–582 (2020).

52. J. A. Backer, D. Klinkenberg, J. Wallinga, Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance.* **25** (2020), doi:10.2807/1560-7917.ES.2020.25.5.2000062.

53. C. McAlloon, Á. Collins, K. Hunt, A. Barber, A. W. Byrne, F. Butler, M. Casey, J. Griffin, E. Lane, D. McEvoy, P. Wall, M. Green, L. O’Grady, S. J. More, Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open.* **10** (2020), doi:10.1136/BMJOPEN-2020-039652.

54. I. Bains, R. Antia, R. Callard, A. J. Yates, Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. *Blood.* **113**, 5480–5487 (2009).

55. C. Rydznski Moderbacher, S. I. Ramirez, J. M. Dan, A. Grifoni, K. M. Hastie, D. Weiskopf, S. Belanger, R. K. Abbott, C. Kim, J. Choi, Y. Kato, E. G. Crotty, C. Kim, S. A. Rawlings, J. Mateus, L. P. V. Tse, A. Frazier, R. Baric, B. Peters, J. Greenbaum, E. Ollmann Saphire, D. M. Smith, A. Sette, S. Crotty, Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell.* **183**, 996–1012.e19 (2020).

56. Q. Qi, Y. Liu, Y. Cheng, J. Glanville, D. Zhang, J. Y. Lee, R. A. Olshen, C. M. Weyand, S. D. Boyd, J. J. Goronzy, Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A.* **111**, 13139–13144 (2014).

57. R. M. Zinkernagel, Lymphocytic choriomeningitis virus and immunology. *Curr Top Microbiol Immunol.* **263**, 1–5 (2002).

Acknowledgments:

5 Funding for COVIDsortium was donated by individuals, charitable Trusts, and corporations, including Goldman Sachs, Kenneth C. Griffin, The Guy Foundation, GW Pharmaceuticals, Kusuma Trust, and Jagclif Charitable Trust, and enabled by Barts Charity with support from UCLH Charity. R.K.G. is funded by the National Institute for Health Research (DRF-2018-11-ST2-004). J.C.M., C.M., and T.A.T. are directly and
10 indirectly supported by the University College London Hospitals NHS Foundation Trust (UCLH) and Barts NIHR Biomedical Research Centers and through the British Heart Foundation (BHF) Accelerator Award (AA/18/6/34223). MM is supported by a Cancer Research UK studentship. T.A.T. is funded by a BHF Intermediate Research Fellowship (FS/19/35/34374). M.N. is supported by the Wellcome Trust (207511/Z/17/Z) and by NIHR Biomedical Research Funding to UCL and UCLH. M.K.M. is supported by
15 UKRI/NIHR UK-CIC and Wellcome Trust (214191/Z/18/Z). G.S.T. is supported by a UK Medical Research Council Clinician Scientist Fellowship (MR/N007727/1). R.J.B. and D.M.A. are supported by UKRI/MRC Newton (MR/S019553/1, MR/R02622X/1, MR/V036939/1, MR/W020610/1), NIHR Imperial Biomedical Research Center (BRC): ITMAT, Cystic Fibrosis Trust SRC (2019SRC015), NIHR EME Fast Track (NIHR134607), NIHR Long Covid (COV-LT2-0027), Innovate UK (SBRI 10008614), and Horizon 2020 Marie Skłodowska-Curie Innovative Training Network (ITN) European Training Network (no. 860325). A.M. is supported by Rosetrees Trust, The John Black Charitable Foundation, and Medical College of St Bartholomew's Hospital Trust. B.M.C. is supported by the Rosetrees Trust, the NIHR UCLH BRC, and the Wellcome Trust. Support was also provided by UK Medical Research Council (T.D, Y.P), 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, S.L.F), National Institutes of Health, National Key R&D Program of China (2020YFE0202400) (T.D), NIHR Oxford Biomedical Research Centre (J.C.K, A.J.M)

Author contributions:

Conceptualization: MN, BC

30 Data acquisition and analysis: MM, YP, CT, MM , GN, SB , AC, JR, MW, XY, GL, SLF, TR

Visualization: CT, MM, TP

Funding acquisition: JCM, MN

Project administration: CM, TT, JCM

35 Supervision: TD, FB, AMcK, CM, TT, MN, BC, EG, SR-Z, CP, JG, AS, TB, DMA, RJB, MKM

Writing – original draft: MM, MN, BC

Writing – review & editing: CP, AS, RB, EG, FB, CT, BC, MM, MN, MKM, AM

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

40 **Data and materials availability:** The COVID_sortium TCR sequencing data are available at NCBI Short Read Archive accession number SUB9362448

(<https://www.ncbi.nlm.nih.gov/sra>). The mouse TCR sequences have been submitted to the Sequence Read Archive under identifier PRJNA771880.

https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA771880&o=acc_s%3Aa All other data are available in the main text or the supplementary materials. All code for analysis and generating individual figure panels are available from BC on request.

5