Distorted TCR repertoires define multisystem inflammatory syndrome in children

Amna Malik* [1], Eszter N. Tóth* [2], Michelle S. Teng [2], Jacob Hurst [2], Eleanor Watt [3], Lauren Wise [4], Natalie Kent [4], Jack Bartram [5], Louis Grandjean [6], Margarita Dominguez-Villar [7], Stuart Adams† [4] and Nichola Cooper† [1]

[1] Centre for Haematology, Department of Immunology and Inflammation, Imperial College London, London, United Kingdom
[2] Etcembly Ltd, Magdalen Centre, Robert Robinson Way, Oxford, United Kingdom
[3] Molecular and Cellular Immunology Department, UCL Great Ormond Street Institute of Child Health, London, United Kingdom
[4] SIHMDS-Haematology, Great Ormond Street Hospital for Children, London, United Kingdom
[5] Department of Haematology, Great Ormond Street Hospital for Children, London, United Kingdom
[6] Paediatric Infectious Diseases, Great Ormond Street Hospital for Children, London, United Kingdom
[7] Department of Infectious Diseases, Imperial College London, London, United Kingdom

These authors have contributed equally to this work

†These authors have jointly supervised

Corresponding author:
Nichola Cooper
Tel: +44 20 3313 4017
E-mail: n.cooper@imperial.ac.uk

Keywords: Sars-Cov2, coronavirus, COVID-19, multisystem inflammatory syndrome, MIS-C, superantigen, TCR, immune repertoire, sequencing

Abstract:

While the majority of children infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) display mild or no symptoms, rare individuals develop severe disease presenting with multisystem inflammatory syndrome (MIS-C). The reason for variable clinical manifestations is not understood. Here, we carried out TCR sequencing and conducted comparative analyses of TCR repertoires between children with severe (n=12) or mild (n=8) COVID-19. We compared these repertoires with unexposed individuals (samples collected pre-COVID-19 pandemic: n=8) and with the Adaptive Biotechnologies MIRA dataset, which includes over 135,000 high-confidence SARS-CoV-2-specific TCRs. We show that the repertoires of severely ill children are characterised by the expansion of TRBV11-2 chains with high junctional and CDR3 diversity. Moreover, the CDR3 sequences of TRBV11-2 clones shift away from SARS-CoV-2 specific T cell clones, resulting in distorted TCR repertoires. In conclusion, our study reports that CDR3-independent expansion of TRBV11-2+ cells, lacking SARS-CoV-2 specificity, defines severity of disease in children.

Introduction

Coronavirus disease 2019, COVID-19, caused by the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is associated with high morbidity and mortality in older individuals, and in those with additional comorbidities. In contrast, children represent a small proportion of COVID-19, comprising less than 2% of cases worldwide [1,2]. While most children with COVID-19 are asymptomatic or present with mild disease, rare individuals develop severe disease presenting with multisystem hyperinflammatory syndrome (MIS-C) including persistent fever, severe abdominal pain, diarrhoea, myocardial dysfunction, cardiogenic shock, rash and neurological disorders [3,4]. This disparity in symptoms is not understood but could represent a difference in the T cell response to SARS-CoV-2.

Whilst it is clear that the adaptive immune response plays an important part in clearance of SARS-CoV-2 infection[5–7], the exact role of T cells in the resolution or potential exacerbation of SARS-CoV-2 infection is not known[8]. A large number of unexposed individuals have SARS-CoV-2 reactive CD4+ memory T cells and these memory T cells have been shown to exhibit cross-reactivity against seasonal “common cold” coronavirus strains [8,9 10 11]. In addition, studies to date have shown that T cell responses develop in almost all patients with confirmed SARS-

CoV-2 infection [12] and remain detectable for several months following infection [8]. In adult COVID-19 patients $CD8^+$ T Cell activation status evolves with disease severity in a non-monotonous way [13]: effector-like cell clusters expand in mild disease and fall during severe disease with the highest level of T-cell polyfunctionality in moderately ill patients. In children activation and proliferation level of CX3CR1+ CD8+ T-cells is much higher in multisystem inflammatory syndrome compared to mild COVID-19 [14]. These cells can interact with fraktalkine-expressing activated endothelium and patrol vasculature, thus this interaction might explain the cardiovascular involvement in children with multisystem inflammatory syndrome. However, whether specific T-cell clones contribute to the hyperinflammatory state or if there is a difference in the T-cell repertoire composition, structure and antigen-specificity-profile of children with mild or severe COVID-19 is still unknown.

T cell receptor sequencing allows the detection and quantification of specific T cell clones and enables us to capture unique patient TCR repertoires. We hypothesized that comparing the TCR repertoire in children with mild or severe COVID-19 and contrasting with either unexposed individuals or COVID specific data sets, could reveal TCR repertoire features that could help understand features associated with severe disease in children.

Results and Discussion

Cohort characteristics (Table 1)

The cohort included children at Great Ormond street Hospital, London, UK with PCR confirmed SARS-CoV-2 defined as having either mild disease (asymptomatic, cough or fever (n=8)) or multisystem hyperinflammatory syndrome (MIS-C (n=12)).

Of the patients tested (n=14), 86% were SARS-CoV-2 antibody positive. Half of the patients (50%, n=10) had pre-existing co-morbidities: 86% (n=7) of the mild disease patients and 25% (n=3) of severe disease patients. The majority of the mild disease patients (n=6, 75%) and 25% (n=3) of severe disease patients were lymphopenic at the point of sampling. The detailed patient information is shown in Table 1.

A control cohort included DNA samples taken from children who had no exposure to SARS-CoV-2 and were collected before the COVID-19 pandemic (n = 8).

This study was done in accordance with The Multi Centre Research Ethics Committee in Wales guidelines MREC Wales reference 06/Q0508/16.

TCR sequencing and repertoire analysis was performed using bulk DNA extracted from blood samples **(Figure 1a).** TCR repertoire metrics can be found in Supplementary Table 1.

The repertoires of severely ill children are characterized by the expansion of TRBV11-2.

We performed principal component analysis (PCA) of TRBV gene usage to determine their global distributions between patient groups with different symptom severity. Patient cohorts showed separation along the first principal component in children with severe disease clustering clearly apart from the children with no or mild symptoms. **(Figure 1b).**

In order to better characterise TRBV gene expression skewing in children with severe disease, we performed differential gene expression analysis of TRBV genes between patient cohorts. This revealed that the expansion of TRBV11-2 chain is mostly responsible for the different TRBV gene usage pattern of children with severe disease **(Figure 1c).** The fraction of TRBV11-2 in a repertoire did not show correlation with the patient's age, the number of days since COVID-19 diagnosis or the antibody status **(Supplementary Figure 1)**.

TRBV11-2 has high junctional diversity in all patient cohorts

To interrogate whether the expansion of TRBV11-2 in the severely ill children cohort is associated with a specific CDR3b motif of J genes usage, we analysed the junctional diversity of TRBV11-2 chains. We compared the frequencies of rearranged J genes in children with mild symptoms to children with severe disease. We did not observe a difference between the cohorts indicating that the expansion of TRBV11-2 in children with severe COVID-19 was not driven by clones harbouring specific TRBV-TRBJ junctions, and that TRBV11-2 had high junctional diversity in all of our patient cohorts **(Figure 1d).** The alignment of expanded TRBV11-2 CDR3 sequences from severely ill patients did not reveal the presence of an enriched CDR3 sequence motif, thus we concluded that the expansion of this V-gene was unrelated to the sequence of the CDR3 peptide binding motif **(Figure 1e)**. This finding is in line with two recent studies suggesting that a superantigen-like sequence motif highly similar to staphylococcal enterotoxin B (SEB) near the S1/S2 cleavage site of the SARS-CoV-2 spike (S) protein can interact with the CDR2 region of TRBV11-2 and may be able to form a ternary

complex with MHCII [15,16]. A superantigen-like interaction bypasses the antigen-specific CDR3 region and involves only the constant CDR2 region, thus all T-cells expressing a given TRBV gene expand regardless of their peptide-MHC specificity. This can lead to the domination of the TCR repertoire by the superantigen-interacting V-gene, while the virus-specific T-cells fail to expand and respond to the infection [17]. In addition, some of the TCRs harbouring the superantigen-interacting V-gene can be autoreactive causing severe autoimmune reaction in the patient. To this date, the presence of autoantibodies in adults suffering from severe COVID-19 disease has been reported [18,19], but little is known about the involvement of autoreactive T-cells in the development of severe disease in adults or multisystem inflammatory syndrome in children. It is possible to hypothesize that if TRBV11-2 TCRs interact with the SARS-CoV-2 spike (S) protein consistent with a superantigen event, this can trigger the expansion and activation of autoreactive TCR clones causing life-threatening complications.

The expansion of TRBV11-2 in severely ill children does not alter the TCR repertoire's overall architecture

We assessed the effect of severe COVID-19 infection on the children's TCR repertoires by computing global T-cell metrics. We found that children infected with SARS-CoV-2 had a lower number of TCR clones than healthy ones, but there was no correlation between the number of clones and symptom severity within the disease cohort indicating that the expansion of TRBV11-2 did not influence the repertoire's richness (**Figure 2a**). The children's age, number of days after COVID-19 diagnosis or presence of co-morbidities did not affect the repertoire's richness either (**Supplementary figure 2a-c**). All repertoires were diverse as shown by the high values of Shannon evenness index (**Figure 2b**). This finding might seem unexpected taken into account that the TRBV11-2 gene was significantly expanded in children with severe COVID-19, however, as shown above (**Figure 1d-e**), the expanded TRBV11-2 chains had high CDR3 sequence diversity, thus they contributed to the repertoire's overall diversity and richness This supports the hypothesis that TRBV11-2 chains might engage in a superantigen-binding interaction with the Sars-Cov2 S-protein regardless of their CDR3 sequence. [15,16].

In order to gain insight into the architecture of patient repertoires we applied a network analysis approach (**Figure 2c**). This method has been used before to decipher the overall

connectivity structure of antibody and TCR repertoires [20–22]. Upon antigen exposure specific T-cells expand, thus they are more likely to be captured when blood samples are collected, and they are easier to detect by TCR sequencing. Homologous TCRs with highly similar CDR3 sequences often recognise the same antigens [23,24], resulting in highly connected TCR sequence similarity networks, with the most similar sequences forming clusters. We analysed connectivity levels of our patient TCR networks by comparing their graph metrics. Usually, high number of nodes and edges in a network, large clusters and high number of connections per node indicate highly connected networks with the presence of many similar TCR sequences. We found that patients in the severely ill group had slightly higher level of network connectivity indicated by the higher number of nodes. The number of network edges, the mean degree or the size of the largest cluster did not change significantly, thus the level of increase in network connectivity was small. This increase might be explained by the presence of more similar clones that arise in response to the more severe disease and prolonged antigen exposure. Overall, the patient repertoires were robust, SARS-CoV-2 infection and the expansion of TRBV11-2 chains in the severely ill cohort did not cause a major change in the immune networks' overall structure (**Figure 2d**).

TRBV11-2 clone sequences of severely ill children shift away from the COVID-19-specific clones

By mapping the clones to the publicly available Adaptive MIRA dataset (described in methods) we can identify TCRs with potential to be SARS-Cov2-specific [25,26]. This database contains more than 130,000 unique TCRβ sequences with known specificities to SARS-CoV-2 antigens from a high number of donors with diverse HLA-backgrounds (MIRA clones). We found that both class I and class II MIRA clones were ubiquitously expressed in our patient repertoires, and the number of identified MIRA clones in a repertoire was directly proportional to the total number of clones (**Supplementary Figure 3**). To assess the overall level of similarity to MIRA clones in a repertoire we defined the *distance to MIRA* measure for each clone (**Figure 3a**). Briefly, a distance matrix was constructed between all clones and all MIRA clones in a repertoire, and the distance to the closest MIRA clone was identified. Naturally, MIRA clones will have a *distance to MIRA* score of 0.

In order to interrogate if the properties of TRBV11-2 clones change in the repertoires of severely ill children compared to children with mild symptoms, we plotted their *distance to MIRA* distributions. In children with mild symptoms, the majority of TRBV11-2 clones have a CDR3 sequence 3-4 amino acids different from a class I MIRA clone (**Figure 3b**). The distribution shifts to higher *distance to MIRA* values in severely ill children, the maximum density being around *distance to MIRA* 4-5 accompanied by a longer tail in the region on high *distance to MIRA* values. We fitted the curves with Gaussian probability distribution functions and achieved significantly better fit when symptom severity was taken into account, indicating that there is a significant difference in the distance to MIRA distributions of TRBV11-2 clones in children with mild and severe disease (**Figure 3b**). All other TRBV chains had the same distance to class I MIRA distribution in mild and severe disease (**Supplementary Figure 4**).

TCRs recognizing the same antigen often have highly similar CDR3 sequences [24,26], therefore, TRBV11-2 clones shifting away from the Sars-Cov2-specific MIRA hits in severely ill children suggests that these clones might be less effective at binding SARS-Cov-2 antigens. In severely ill children, TRBV11-2 chain is highly expanded with no CDR3 motif expansion or specific J-gene usage and the CDR3 sequences of these expanded TRBV11-2 clones shift away from the SARS-CoV-2-specific MIRA hits. This suggests that the expansion of TRBV11-2 clones might be independent of the classical CDR3-peptide-MHC-mediated antigen recognition. As SARS-CoV-2 Spike protein has been suggested to have a superantigen structure[15,16] and considering our findings, it is likely that the observed TCR repertoire skewing is superantigen-induced which results in the repertoire being dominated by the expression of non-specific TCRs that are unable to respond to the infection and instead contributing to the hyperinflammatory state.
There are some striking clinical similarities between MIS-C and toxic shock syndrome (TSS), caused by bacterial superantigens [27,28]. Superantigens simultaneously bind major histocompatibility complex (MHC) class II (MHCII) molecules on antigen presenting cells and T cell receptors (TCRs) of both CD4+ and CD8+ T cells [29]. They are able to circumvent TCR specificity by binding to specific TCR β-chains in a complementary-determining region 3 (CDR3)-independent manner, resulting in broad T cell activation. In patients with MIS-C, skewing of specific TCR β Variable (V) genes, with diverse CDR3 and Joining (J) usage, has been

reported to correlate with disease severity, consistent with superantigen triggered immune activation [16].

The distance distribution of TRBV11-2 clones to class II MIRA hits showed a slight, but not significant shift to higher values (**Figure 3c**). Although the shift in the median values is evident, the high variation in the data didn't allow us to accurately estimate the distribution curves, thus our model did not determine significant difference between mild and severe patients. The distance to class II MIRA distributions of all other TRBV genes did not show any significant difference between our patient cohorts either (**Supplementary Figure 5**). The high variation in the data can be explained by the low number of class II MIRA hits in our repertoires. Class II MIRA clones accounted for about 0.4% of all clones in a repertoire (**Supplementary Figure 3c-d**), whereas about 5% of each repertoire was a class I MIRA clone (**Supplementary Figure 3a-b**). This discrepancy can be explained by the fact that the majority of the Adaptive MIRA database consists of class I hits. The MIRA database contains several data releases with most of them reporting TCRs that interact with antigens in the context of MHC class I.

In summary, the T cell repertoire of severely ill children infected with SARS-CoV-2 is distorted, by TRBV11-2+ T cell clonal expansion and activation, which could be super antigen induced causing an aberrant immune response and leading to clinical manifestations reminiscent of toxic shock syndrome [27,28]. We report the use of two metrics to define the severity of disease in children infected with SARS-CoV-2: 1) CDR3-independent expansion of TRBV11-2+ T cells, 2) a lack of SARS-CoV-2 specificity in TRBV11-2+ T cells, measured by distance to Sars-Cov2-specific MIRA clones. These two metrics can serve as biomarkers for early detection of multisystem inflammatory syndrome in children (MIS-C) guiding physicians to start precision immunotherapeutics that can prevent the development of severe, life-threatening complications and lasting disability in children.

Conflict of interest statement

N.C and M.D.V. are partially funded by Imperial College NIHR BRC. A.M. is funded by the Jon Moulton Charity. E.N.T., M.S.T. and J.H. have a financial interest in Etcembly Ltd.

Methods

TCR sequencing

Next generation sequencing of the T-cell receptor (TCR) was carried out as previously described [30]. Briefly, DNA was extracted from patient blood samples using DNeasy Blood & Tissue kit (Qiagen), quantified using a Qubit Fluorometer (ThermoFisher Scientific) and amplified by multiplex-PCR of rearranged variable, diverse, joining (VDJ) segments of the TCR genes, which encode the hypervariable CDR3 domain. The products were size selected using Pronex beads (Promega) and subsequently sequenced on a MiSeq (Illumina).

Analysis of the raw TCR sequences was performed using MiXCR[31]. A built-in library of reference germline V, D, J, and C gene loci from the ImMunoGeneTics (IMGT) database (imgt.org) is employed by MiXCR. The IMGT nomenclature for TCR gene segments is used throughout the study.

MIRA data set

Adaptive Biotechnologies have created a MIRA dataset of T cell clones, which includes over 135,000 high-confidence SARS-CoV-2-specific TCRs (MIRA clones) at the time of this study [25,26]. The MIRA assay is a high-throughput multiplex tool that maps TCRs binding to SARS-Cov2 virus epitopes by exposing PBMCs to Sars-Cov2 minigenes or peptide pools, sorting T-cells based on surface expression of activation markers and sequencing the TCRs expressed by these activated T-cells. The dataset is a collection of matching antigen-TCR data from both Sars-Cov2-convalescent subjects and unexposed individuals. The antigens include Sars-Cov2 epitopes presented by a diverse set of MHC class I as well as MHC class II alleles, thus capturing response by CD8+ and CD4+ T-cells

Data analysis and visualization

All visualization and standard statistical analysis were conducted using R version 4.0.3 and Python 3.9.1. Plots were generated using the ggplot2 R package [32]. Correlation was quantified by Spearman's rank correlation coefficient (ρ). Paired analyses were performed by non-parametric paired Wilcoxon test. All tests were performed two-sided with a nominal

significance threshold of $P < 0.05$. In all cases of multiple comparisons, false discovery rate (FDR) correction was performed using the Benjamini-Hochberg procedure. Illustrations were prepared with the BioRender package.

TCR repertoire analysis

TRBV gene usage, TRBV-TRBJ junction frequencies and repertoire global metrics were calculated for each sample. When samples from multiple time points were available for a patient, the mean values were used for downstream analysis. Differential gene expression analysis of TRBV genes was conducted using the EdgeR package [33].

The normalized repertoire richness (R) was calculated as follows:

$$R = \frac{n}{r} * 10^6$$

Where n is the number of clonotypes, and r is the number of reads in a repertoire.

The Shannon evenness index (J') was calculated as in (Attaf et al, Front Immunol, 2018):

$$J' = \frac{H'}{\ln(n)} \qquad H' = \sum_{i=1}^{n} p_i \ln(p_i)$$

Where p_i is the frequency of the i^{th} clonotype in a population of n clonotypes. J' is undefined for monoclonal samples. Low J' values approaching 0 indicate minimal evenness such as after clonal expansion of antigen-specific clones. The maximum value of J is 1, when all clonotypes have equal frequencies, thus the population is perfectly even.

Network analysis

First, multiple samples from the same patient collected at different time points were combined. Next, to adjust for different sequencing coverage and clonal depth among patients we randomly down sampled each repertoire to 1,000 clones. Next, we calculated pairwise amino acid sequence similarity of these 1,000 clones by constructing Levenshtein distance (LD) matrices of their CDR3 sequences using the stringdist package [34]. Networks were generated using the igraph package[35]. Each node in the TCR similarity network represents a

unique amino acid clone, and edges between nodes are constructed by connecting nodes that differ by no more than 2 amino acids (LD1-2) in their CDR3 sequences. The final similarity network contains only nodes that make at least one connection to another node in the network. Clusters were defined as groups of interconnected nodes. Classic graph analysis metrics [36], such as the number of nodes, edges, node degree and cluster sizes were calculated using the igraph package as well. The random repertoire down sampling, network construction and calculation of network metrics were repeated 10 times for each patient and the mean values were used for the downstream analysis.

Distance to MIRA analysis

First, samples from the same patients taken at different time points were combined, and MIRA clones in the repertoires were found by mapping the CDR3 sequences to the Adaptive MIRA database [25,26]. Next, Levenshtein distance matrices (LD) were constructed between all clones and all MIRA clones in a repertoire using the stringdist package [34]. The *distance to MIRA* value was defined as the shortest distance a MIRA clone. Gene-wise *distance to MIRA* distributions were analysed for those TRBV genes that occupied at least 1% of mean expressed repertoire in each patient cohort. These distributions were fitted with Gaussian probability density functions and the difference between patient cohorts was assessed by comparing a model fitting all symptom groups together to a model taking into account the differences between symptom groups. The difference between the two models was determined with ANOVA analysis (as described in [37]), with all p-values corrected for multiple hypothesis testing using Benjamini-Hochberg adjustment.

Figure Legends

Figure 1: TRBV11-2 is expanded in the repertoire of severely ill children a) Overview of study design b) Principal component analysis (PCA) of differential TRBV usage in severely ill children (n=12) children with mild symptoms (n=9), and pre-COVID-19 children (n=8). Values in brackets show the percentage of variation explained by each principal component) Frequencies of TRBV genes in the patient cohorts. Bars indicate mean + SEM. Insert: TRBV11-2 usage in patient cohorts. Differential gene expression analysis of TRBV genes was conducted using the EdgeR package. TRBV11-2 is expanded in severely ill children compared to children with mild symptom (p-val < 0.0001) d) Differential usage of J genes rearranged with TRBV11-2. Bars indicate mean + SEM. TRBV11-2 rearrangement frequencies were analysed using the EdgeR package e) CDR3 diversity of TRBV11-2 in children with mild and severe disease displayed as positional weight matrix. CDR3 sequences containing 13 amino acids are shown as examples.

Figure 2: The expansion of TRBV11-2 in severely ill children doesn't alter the TCR repertoire's overall architecture a-b) TCR repertoire metrics of patients. Significance determined by unpaired Wilcoxon test between each pediatric group, with adjustment for multiple comparisons using Benjamini-Hochberg correction, indicated by: * p<0.05, ** p<0.01, and *** p<0.001. Lack of notation for specified comparisons indicates no statistical significance. c) Overview of sequence similarity network analysis: Each repertoire was downsampled to 1,000 clones. Pairwise amino acid sequence similarity of these 1,000 clones was calculated by constructing Levenshtein distance (LD) matrices of their CDR3 sequences. Each node in the TCR similarity network represents a unique amino acid clone, and edges between nodes are constructed by connecting nodes that differ by no more than 2 amino acids (LD1-2) in their CDR3 sequences. Classic graph metrics, such as node and edge number, node degree and the size of the largest cluster were calculated. The random down sampling and network construction were repeated 10 times and the mean value of the network metrics was used for the downstream analysis c) TCR sequence similarity network metrics of patient cohorts. Significance determined by unpaired Wilcoxon test between each paediatric group, with adjustment for multiple comparisons using Benjamini-Hochberg correction, indicated by: * p<0.05, ** p<0.01, and *** p<0.001. Lack of notation for specified comparisons indicates no statistical significance.

Figure 3: Distance to MIRA analysis a) Overview of distance to MIRA analysis: A distance matrix was constructed between all clones in a repertoire and those that mapped to the Adaptive MIRA database, and the distance to the closest MIRA clone was identified. b) Distance to class I MIRA distribution of TRBV11-2 clones in the repertoires of children with mild or severe COVID-19.c) Distance to class II MIRA distribution of TRBV11-2 clones in the repertoires of children with mild or severe COVID-19.

Distance to class I and class II MIRA distributions were fitted with Gaussian probability density functions and the difference between patient cohorts was assessed by comparing a model fitting all symptom groups together to a model taking into account the differences between symptom groups. The difference between the two models was determined with ANOVA analysis, with all p-values corrected for multiple hypothesis testing using Benjamini-Hochberg adjustment. Solid curves show the probability distribution functions determined by fitting the patient cohorts separately, shaded areas show 75% confidence intervals. P-values denote the statistical assessment of the models fitting all patient cohorts together versus fitting each patient cohort separately. See the distance to MIRA distributions of clones with TRBV gene other than TRBV11-2 in Supplementary Figures 4 and 5.

References

1. Wu, Z. & McGoogan, J. M. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* **323**, 1239 (2020).
2. Feng, Z., Bao, Y., Yang, Y., Zheng, Y. & Shen, K. Severe acute respiratory syndrome coronavirus 2-induced multisystem inflammatory syndrome in children. *Pediatr. Investig.* **4**, 257–262 (2020).
3. Neeland, M. R. *et al.* Innate cell profiles during the acute and convalescent phase of SARS-CoV-2 infection in children. *Nat. Commun.* **12**, 1084 (2021).
4. Jin, Y.-H. *et al.* A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). *Mil. Med. Res.* **7**, 4 (2020).
5. Kuri-Cervantes, L. *et al.* Comprehensive mapping of immune perturbations associated with severe COVID-19. *Sci. Immunol.* **5**, (2020).
6. Wilk, A. J. *et al.* A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nat. Med.* **26**, 1070–1076 (2020).
7. Liao, M. *et al.* Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* **26**, 842–844 (2020).
8. 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.e15 (2020).
9. Braun, J. *et al.* SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* **587**, 270–274 (2020).

10. Ng, O.-W. *et al.* Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* **34**, 2008–2014 (2016).
11. Zhao, J., Zhao, J. & Perlman, S. T Cell Responses Are Required for Protection from Clinical Disease and for Virus Clearance in Severe Acute Respiratory Syndrome Coronavirus-Infected Mice. *J. Virol.* **84**, 9318–9325 (2010).
12. Zuo, J. *et al.* Robust SARS-CoV-2-specific T cell immunity is maintained at 6 months following primary infection. *Nat. Immunol.* (2021) doi:10.1038/s41590-021-00902-8.
13. Su, Y. *et al.* Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19. *Cell* **183**, 1479-1495.e20 (2020).
14. Vella, L. A. *et al.* Deep immune profiling of MIS-C demonstrates marked but transient immune activation compared to adult and pediatric COVID-19. *Sci. Immunol.* **6**, eabf7570 (2021).
15. Cheng, M. H. *et al.* Superantigenic character of an insert unique to SARS-CoV-2 spike supported by skewed TCR repertoire in patients with hyperinflammation. *PNAS* **117**, 25254–25262.
16. Porritt, R. A. *et al.* Identification of a unique TCR repertoire, consistent with a superantigen selection process in Children with Multi-system Inflammatory Syndrome. *bioRxiv* (2020).
17. Fraser, J. D. & Proft, T. The bacterial superantigen and superantigen-like proteins. *Immunol. Rev.* **225**, 226–243 (2008).
18. Bastard, P. *et al.* Autoantibodies against type I IFNs in patients with life-threatening COVID-19. *Science* **370**, eabd4585 (2020).

19. Wang, E. Y. *et al. Diverse Functional Autoantibodies in Patients with COVID-19*. http://medrxiv.org/lookup/doi/10.1101/2020.12.10.20247205 (2020) doi:10.1101/2020.12.10.20247205.
20. Miho, E., Roškar, R., Greiff, V. & Reddy, S. T. Large-scale network analysis reveals the sequence space architecture of antibody repertoires. *Nat. Commun.* **10**, 1–11 (2019).
21. Madi, A. *et al.* T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *eLife* **6**, 1–17 (2017).
22. Pogorelyy, M. V. & Shugay, M. A framework for annotation of antigen specificities in high-throughput T-cell repertoire sequencing studies. *Front. Immunol.* **10**, 1–9 (2019).
23. Dash, P. *et al.* Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* **547**, 89–93 (2017).
24. Neller, M. A. *et al.* Naive CD8+ T-cell precursors display structured TCR repertoires and composite antigen-driven selection dynamics. *Immunol. Cell Biol.* **93**, 625–633 (2015).
25. Nolan, S. *et al.* A large-scale database of T-cell receptor beta (TCRβ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. *Res. Sq.* 1–28 (2020).
26. Snyder, T. M. *et al.* Magnitude and Dynamics of the T-Cell Response to SARS-CoV-2 Infection at Both Individual and Population Levels. *medRxiv* (2020).
27. Riphagen, S., Gomez, X., Gonzalez-Martinez, C., Wilkinson, N. & Theocharis, P. Hyperinflammatory shock in children during COVID-19 pandemic. *The Lancet* **395**, 1607–1608 (2020).
28. Noval Rivas, M., Porritt, R. A., Cheng, M. H., Bahar, I. & Arditi, M. COVID-19–associated multisystem inflammatory syndrome in children (MIS-C): A novel disease that mimics

toxic shock syndrome—the superantigen hypothesis. *J. Allergy Clin. Immunol.* **147**, 57–59 (2021).

29. Li, H., Llera, A., Malchiodi, E. L. & Mariuzza, R. A. The Structural Basis of T Cell Activation by Superantigens. *Annu. Rev. Immunol.* **17**, 435–466 (1999).
30. Bartram, J. *et al.* Accurate Sample Assignment in a Multiplexed, Ultrasensitive, High-Throughput Sequencing Assay for Minimal Residual Disease. *J. Mol. Diagn.* **18**, 494–506 (2016).
31. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
32. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (New York: Springer, 2016).
33. Robinson, M. D., Mccarthy, D. J. & Smyth, G. K. edgeR : a Bioconductor package for differential expression analysis of digital gene expression data. **26**, 139–140 (2010).
34. van der Loo, M. P. J. The stringdist package for approximate string matching. *R J.* **6**, 111–122 (2014).
35. Csardi, G. & Nepusz, T. The igraph software package for complex network research. *InterJournal Complex Syst.* **Complex Sy**, 1695 (2006).
36. Lee, S. H., Kim, P. J. & Jeong, H. Statistical properties of sampled networks. *Phys. Rev. E - Stat. Nonlinear Soft Matter Phys.* **73**, 1–7 (2006).
37. *Nonlinear Regression with R*. (Springer New York, 2009). doi:10.1007/978-0-387-09616-2.

Table 1: Cohort characteristics of SARS-CoV-2 infected children and unexposed children

	Unexposed (n=8)	Mild (n=8)	Severe (n=12)
Age in months (IQR)	**84(120)**	**96 (87)**	**60 (51)**
SARS-CoV-2 PCR Positivity	**N/A**	**100%(2/2)**	**100%(12/12)**
Antibody Test positivity	**N/A**	**86%(12/14)**	**83%(10/12)**
Longitudinal samples	**N/A**	**63%(5/8)**	**8%(1/12)**
Symptoms			
Fever	N/A	5%(1/12)	0
Cough	N/A	5%(1/12)	0
CNS Disease	N/A	0	42%(5/12)
Hyperinflammatory Syndrome	N/A	0	58%(7/12)
None	N/A	75%(6/8)	0
Other co-morbidities			
Dravet Syndrome	NA	5%(1/8)	0
hepatoblastoma	NA	0	5%(1/12)
Neuroblastoma	NA	5%(1/8)	0
TAPDV	NA	0	5%(1/12)
Tonsillitis	NA	0	5%(1/12)
Atypical teratoid rhabdoid tumour	NA	5%(1/8)	0
PNET metastatic tumour	NA	5%(1/8)	0
B cell ALL	NA	10%(2/8)	0
Complex metabolic disorder	NA	5%(1/8)	0
None	100%(10/10)	13%(1/8)	70%(9/12)

Figure 1

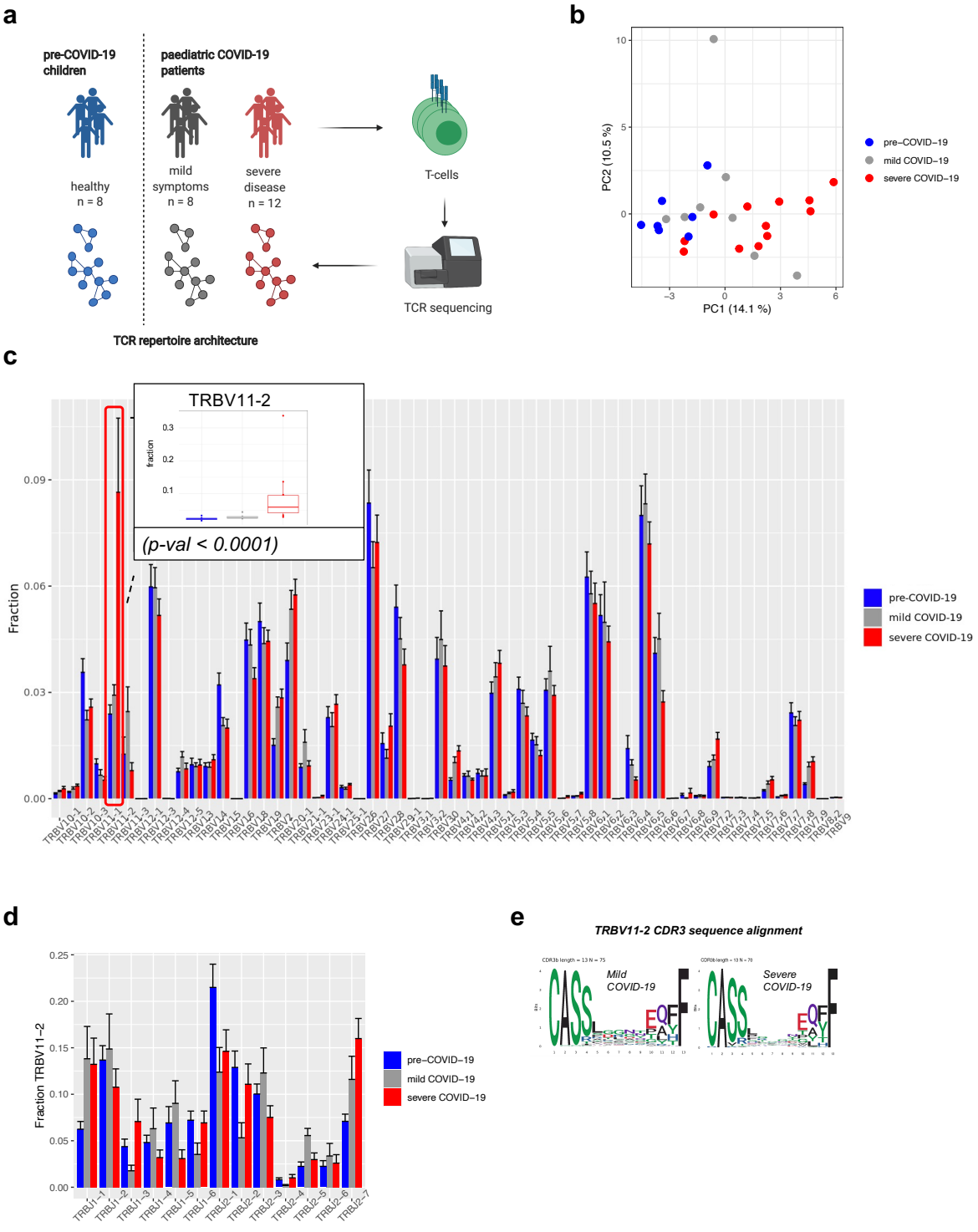

a
pre-COVID-19 children
paediatric COVID-19 patients
healthy n = 8
mild symptoms n = 8
severe disease n = 12
T-cells
TCR sequencing
TCR repertoire architecture
b
PC2 (10.5 %)
PC1 (14.1 %)
pre−COVID−19
mild COVID−19
severe COVID−19
c
TRBV11-2
fraction
(p-val < 0.0001)
Fraction
pre-COVID-19
mild COVID-19
severe COVID-19
d
Fraction TRBV11−2
pre−COVID−19
mild COVID−19
severe COVID−19
e
TRBV11-2 CDR3 sequence alignment
Mild COVID-19
Severe COVID-19

Figure 2

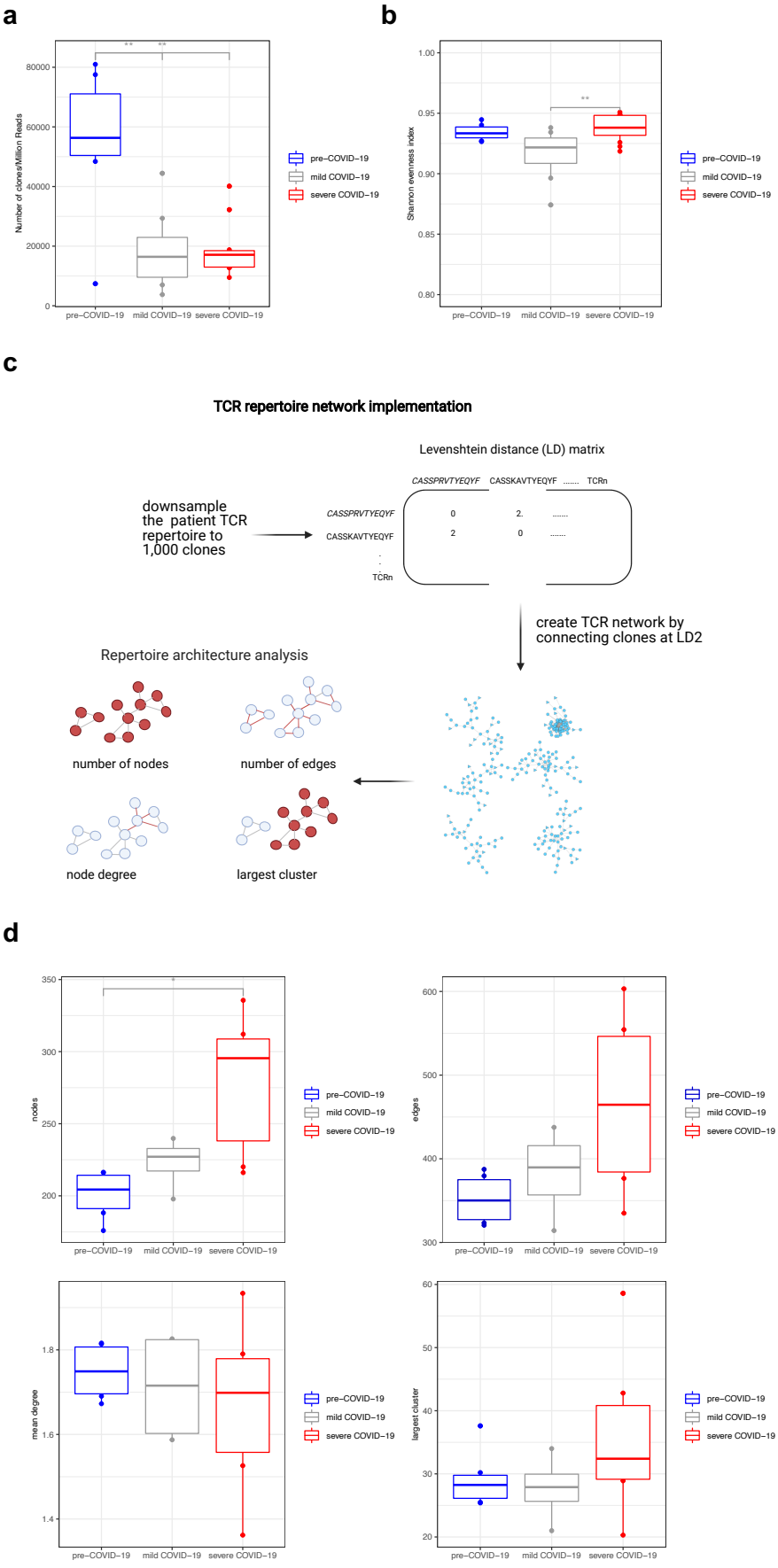
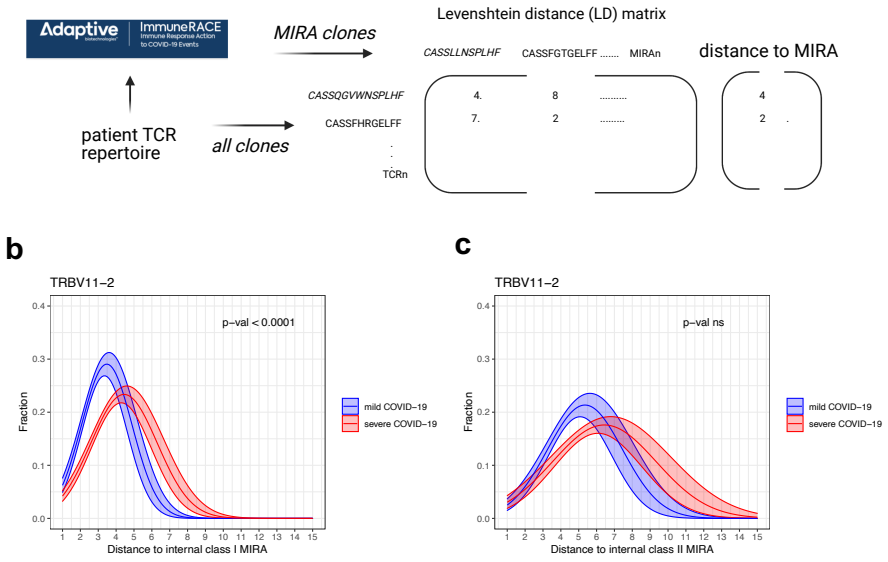

a
b
c
d
Number of clones/Million Reads
Shannon evenness index
pre−COVID−19
mild COVID−19
severe COVID−19
TCR repertoire network implementation
Levenshtein distance (LD) matrix
downsample the patient TCR repertoire to 1,000 clones
create TCR network by connecting clones at LD2
Repertoire architecture analysis
number of nodes
number of edges
node degree
largest cluster
nodes
edges
mean degree
largest cluster

Figure 3


a
Adaptive
ImmuneRACE
Immune Response Action to COVID-19 Events
patient TCR repertoire
MIRA clones
all clones
CASSLLNSPLHF
CASSQGVWNSPLHF
CASSFHRGELFF
TCRn
Levenshtein distance (LD) matrix
CASSFGTGELFF ..... MIRAn
4.
7.
8
2
4
2
distance to MIRA
b
TRBV11–2
p–val < 0.0001
Fraction
Distance to internal class I MIRA
mild COVID–19
severe COVID–19
c
TRBV11–2
p–val ns
Fraction
Distance to internal class II MIRA
mild COVID–19
severe COVID–19