Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T cell memory formation after mild COVID-19 infection

Anastasia A. Minervina[1], Ekaterina A. Komech[1,2], Aleksei Titov[3], Meriem Bensouda Koraichi[4], Elisa Rosati[5], Ilgar Z. Mamedov[1,2,6,7], Andre Franke[5], Grigory A. Efimov[3], Dmitriy M. Chudakov[1,2,6,8], Thierry Mora[4], Aleksandra M. Walczak[4], Yuri B. Lebedev[1,9], Mikhail V. Pogorelyy[1,2]

[1] *Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russia*
[2] *Pirogov Russian National Research Medical University, Moscow, Russia*
[3] *National Research Center for Hematology, Moscow, Russia*
[4] *Laboratoire de physique de l'École normale supérieure, ENS, PSL, Sorbonne Université, Université de Paris, and CNRS, Paris, France*
[5] *Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany*
[6] *Masaryk University, Central European Institute of Technology, Brno, Czech Republic*
[7] *V.I. Kulakov National Medical Research Center for Obstetrics, Gynecology and Perinatology, Moscow, Russia*
[8] *Center of Life Sciences, Skoltech, Moscow, Russia*
[9] *Moscow State University, Moscow, Russia*

Abstract. COVID-19 is a global pandemic caused by the SARS-CoV-2 coronavirus. T cell response is a critical part of both individual and herd immunity to SARS-CoV-2 and the efficacy of developed vaccines. However neither the dynamics and cross-reactivity of the SARS-CoV-2-specific T cell response nor the diversity of resulting immune memory are well understood. In this study we use longitudinal high-throughput T cell receptor sequencing to track changes in the T cell repertoire following two mild cases of COVID-19 infection. In both donors we identified SARS-CoV-2-responding CD4+ and CD8+ T cell clones. We describe characteristic motifs in TCR sequences of COVID-19-reactive clones, suggesting the existence of immunodominant epitopes. We show that in both donors the majority of infection-reactive clonotypes acquire memory phenotypes. Certain CD4+ clones were detected in the memory fraction at the pre-infection timepoint, suggesting participation of pre-existing cross-reactive memory T cells in the immune response to SARS-CoV-2.

INTRODUCTION

COVID-19 is a global pandemic caused by the novel SARS-CoV-2 betacoronavirus [1]. T cells are crucial for clearing respiratory viral infections and providing long-term immune memory [2, 3]. Two major subsets of T cells participate in the immune response to viral infection in different ways: activated CD8+ T cells directly kill infected cells, while subpopulations of CD4+ T cells produce signaling molecules that regulate myeloid cell behaviour, drive and support CD8 response and the formation of long-term CD8 memory, and participate in the selection and affinity maturation of antigen specific B-cells, ultimately leading to the generation of neutralizing antibodies. In SARS-1 survivors, antigen-specific memory T cells were detected up to 11 years after the initial infection, when viral-specific antibodies were undetectable [4, 5]. The T cell response was shown to be critical for protection in SARS-1-infected mice [6]. Patients with X-linked agammaglobulinemia, a genetic disorder associated with lack of B cells, have been reported to recover from symptomatic COVID-19 [7, 8], suggesting that in some cases T cells are sufficient for viral clearance. Theravajan et al. showed that activated CD8+HLA-DR+CD38+ T cells in a mild case of COVID-19 significantly expand following symptom onset, reaching their peak frequency of 12% of CD8+ T cells on day 9 after symptom onset, and contract thereafter [9]. Given the average time of 5 days from infection to the onset of symptoms [10], the dynamics and magnitude of T cell response to SARS-CoV-2 is similar to that observed after immunization with live vaccines [11]. The exact immunodominant CD8+ and CD4+ SARS-CoV-2 epitopes are yet unknown. However, SARS-CoV-2-specific T cells were detected in COVID-19 survivors by activation following stimulation with SARS-CoV-2 proteins [12], or by viral protein-derived peptide pools [13, 14]. Some of the T cells activated by peptide stimulation were shown to have a memory phenotype [13], and some potentially cross-reactive CD4+ T cells were found in healthy donors [14, 15].

T cells recognise short pathogen-derived peptides presented on the cell surface of the Major Histocompatibility Complex (MHC) using hypervariable T cell receptors

(TCR). TCR repertoire sequencing allows for the quantitative tracking of T cell clones in time, as they go through the expansion and contraction phases of the response. It was previously shown that quantitative longitudinal TCR sequencing is able to identify antigen-specific expanding and contracting T cells in response to yellow fever vaccination with high sensitivity and specificity [16–18]. Not only clonal expansion but also significant contraction from the peak of the response are distinctive traits of T cell clones specific to the virus [17].

In this study we use longitudinal TCRalpha and TCRbeta repertoire sequencing to quantitatively track T cell clones that significantly expand and contract after a mild COVID-19, and determine their phenotype. We reveal the dynamics and the phenotype of the memory cells formed after infection, identify pre-existing T cell memory clones participating in the response, and describe public TCR sequence motifs of SARS-nCoV-2-reactive clones, suggesting a response to immunodominant epitopes.

RESULTS

In the middle of March (day 0) donor W female and donor M (male, both healthy young adults), returned to their home country from the one of the centers of the COVID-19 outbreak in Europe at the time. Upon arrival, according to local regulations, they were put into strict self-quarantine for 14 days. On day 3 of self-isolation both developed low grade fever, fatigue and myalgia, which lasted 4 days and was followed by a temporary loss of smell for donor M. On days 15, 30, 37 and 45 we collected peripheral blood samples from both donors. The presence of SARS-CoV-2 specific antibodies in the plasma was measured at all timepoints using SARS-CoV-2 S-RBD domain specific ELISA (Fig. S1). From each blood sample we isolated PBMCs (peripheral blood mononuclear cells, in two biological replicates), CD4+, and CD8+ T cells. Additionally, on days 30 and 45 we isolated four T cell memory subpopulations (Fig. S2): Effector Memory (EM: CCR7-CD45RA-), Effector Memory with CD45RA re-expression (EMRA: CCR7-CD45RA+), Central Memory (CM: CCR7+CD45RA-), and Stem Cell-like Memory (SCM: CCR7+CD45RA+CD95+). From all samples we isolated RNA and performed TCRalpha and TCRbeta repertoire sequencing as previously described [19]. For both donors, TCRalpha and TCRbeta repertoires were obtained for other projects one and two years prior to infection. Additionally, TCR repertoires of multiple samples for donor M – including sorted memory subpopulations – are available from a published longitudinal TCR sequencing study after yellow fever vaccination (donor M1 samples in [16]).

From previously described activated T cell dynamics for SARS-CoV-2 [9], and immunization with live vaccines [11], the peak of the T cell expansion is expected around day 15 post-infection, and responding T cells significantly contract by day 45. However, Weiskopf et al. [13] reports an increase of SARS-CoV-2-reactive T cells at later timepoints, peaking in some donors after 30 days following symptom onset. To identify groups of T cell clones with similar dynamics in an unbiased way, we used Principal Component Analysis (PCA) in the space of T cell clonal trajectories (Fig. 1b and c). In both donors, and in both TCRalpha and TCRbeta repertoires, we identified three clusters of clones with distinct dynamics. The first cluster (Fig. 1bc, purple) corresponded to abundant TCR clonotypes which had constant concentrations across timepoints, the second cluster (Fig. 1bc, green) showed contraction dynamics from day 15 to day 45, and the third cluster (Fig. 1bc, yellow), showed an unexpected clonal expansion from day 15 with a peak on day 37 followed by contraction. The clustering and dynamics are similar in both donors and are reproduced in TCRbeta (Fig. 1bc) and TCRalpha (Fig. S3) repertoires. We next used edgeR, a software for differential gene expression analysis [20], to specifically detect changes in clonotype concentration between pairs of timepoints in a statistically reliable way. EdgeR uses biological replicate samples collected at each timepoint to train a negative-binomial noise model. We identified 374 TCRalpha and 373 TCRbeta clonotypes in donor W, and 797 TCRalpha and 672 TCRbeta in donor M significantly contracted from day 15 to day 45 (largely overlapping with cluster 2 of clonal trajectories). 438 TCRalpha and 533 TCRbeta for donor W, and 531 TCRalpha and 599 TCRbeta clonotypes for donor M were significantly expanded from day 15 to 37 (corresponding to cluster 3 of clonal trajectories). Reproducing the analysis using NoiseET, a Bayesian differential expansion model [21], yielded similar results (see Fig. S4), suggesting that our results are robust to the choice of statistical model and parameter inference procedure. Note that, to identify putatively SARS-CoV-2 reactive clones, we only used post-infection timepoints, so that our analysis can be reproduced in other patients and studies where pre-infection timepoints are unavailable. However, tracking the identified responding clones back to pre-infection timepoints reveals strong clonal expansions from pre- to post-infection (Fig. 1de, Fig. S3cd). For brevity, we further refer to clonotypes significantly contracted from day 15 to 45 as *contracting* clones and clonotypes significantly expanding from day 15 to 37 as *expanding* clones. For each contracting and expanding clone we determined their CD4/CD8 phenotype using separately sequenced repertoires of CD4+ and CD8+ subpopulations (see Methods). Both CD4+ and CD8+ subsets participated actively in the response (Fig. 1de). Interestingly, clonotypes expanding after day 15 were significantly biased towards the CD4+ phenotype, while contracting clones had balanced CD4/CD8 phenotype fractions in both donors (Fisher exact test, $p < 0.0001$ for both donors).

On days 30 and 45 we identified both contracting (Fig.

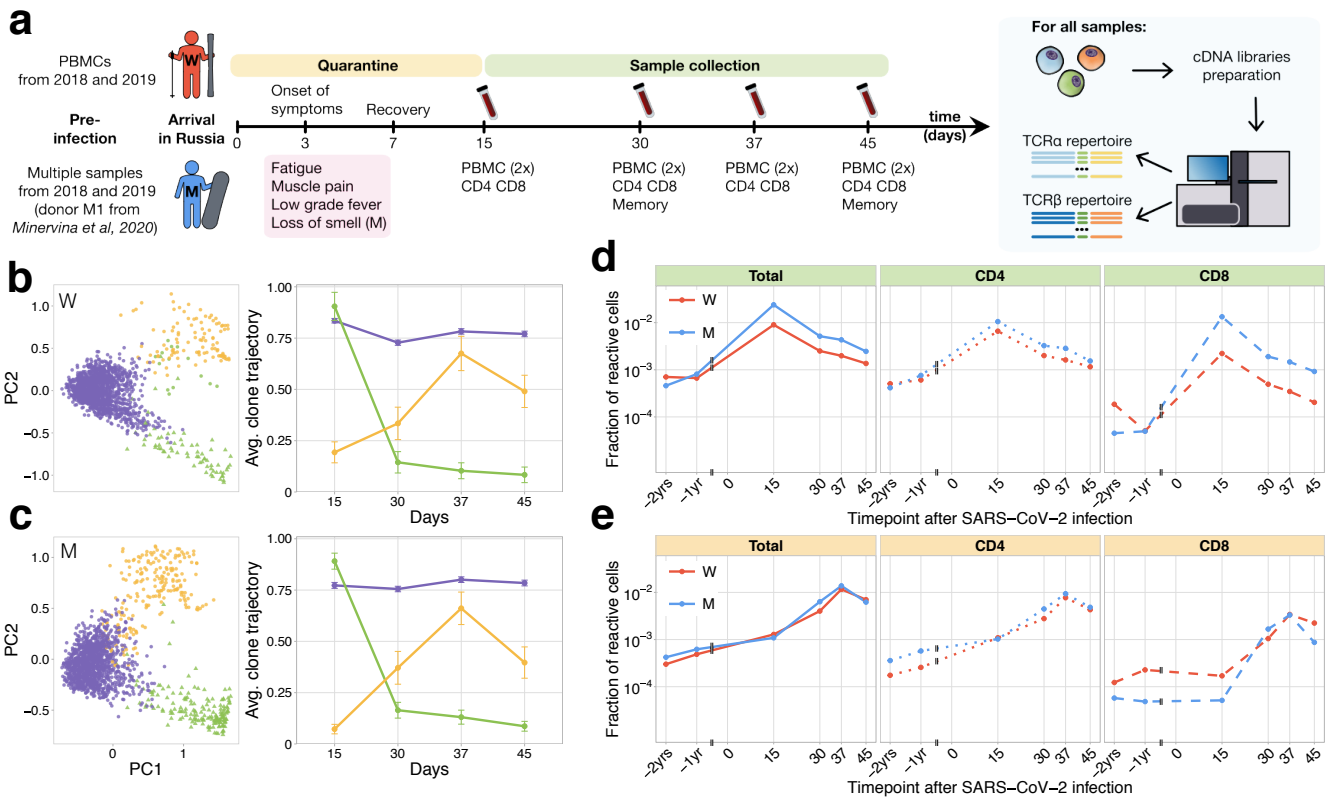

FIG. 1. **Longitudinal tracking of T cell clones after mild COVID-19. a, Study design.** Peripheral blood of two donors was sampled longitudinally on days 15, 30, 37, 45 after arrival in Russia. At each timepoint, we evaluated SARS-CoV-2-specific antibodies using ELISA (Fig. S1) and isolated PBMCs in two biological replicates. Additionally, CD4+ and CD8+ cells were isolated from a separate portion of blood, and EM, CM, EMRA, SCM memory subpopulations were FACS sorted on days 30 and 45. For each sample we sequenced TCRalpha and TCRbeta repertoires. For both donors pre-infection PBMC repertoires were sampled in 2018 and 2019 for other projects. **b,c, PCA of clonal temporal trajectories identifies three groups of clones with distinctive behaviours.** Left: first two principal components of the 1000 most abundant TCRbeta clonotype normalized frequencies in PBMC at post-infection timepoints. Right: mean ± 2.96 SE of the normalized trajectory of TCR clones in each cluster. **Contracting (d) and expanding (e) clones include both CD4+ and CD8+ T cells, and are less abundant in pre-infection repertoires.** T cell clones significantly contracted from day 15 to day 45 **(d)** and significantly expanded from day 15 to day 37 **(e)** were identified using edgeR in both donors. The fraction of contracting **(d)** and expanding **(e)** TCRbeta clonotypes in the total repertoire (corresponding to the fraction of responding cells of all T cells) is plotted in log-scale for all reactive clones (left), reactive clones with the CD4 (middle) and CD8 (right) phenotypes. Similar dynamics were observed in TCRalpha repertoires (Fig. S3), and for significantly expanded/contracted clones identified with the NoisET Bayesian differential expansion statistical model (Fig. S4).

2a,b) and expanding (SI Fig. 5a-c) T cell clones in the memory subpopulations of peripheral blood. Both CD4+ and CD8+ responding clones were found in the CM and EM subsets, however CD4+ were more biased towards CM, and CD8+ clones much more represented in the EMRA subset. A small number of both CD4+ and CD8+ responding clonotypes were also identified in the SCM subpopulation, which was previously shown to be a long-lived T cell memory subset [22]. Intriguingly, a number of responding CD4+ clones, but very few CD8+ clones, were also represented in the repertoires of both donors 1 and 2 years before the infection. Pre-existing clones were expanded after infection, and contracted afterwards for both donors (SI Fig. 6). For donor M, for whom we had previously sequenced memory subpopulations before the infection [16], we were able to identify pre-existing SARS-CoV-2-reactive CD4+ clones in the CM subpopulation 1 year before the infection. Interestingly, on day 30 after infection the majority of pre-infection CM clones were detected in the EM subpopulation, suggesting recent T cell activation and a switch of the phenotype from memory to effector. These clones might represent memory T cells cross-reactive for other infections, e.g. other human coronaviruses. The search for responding clones TCR-beta amino acid sequences in VDJdb [23] – a database of TCRs with known specificity – resulted in essentially no overlap: a total of 3 matches all corresponded to the EBV (Epstein-Barr virus) epitope presented by the HLA-A*03 MHC allele, which is absent in both donors (SI Table 1). At the time of the analysis there were no SARS-1, SARS-

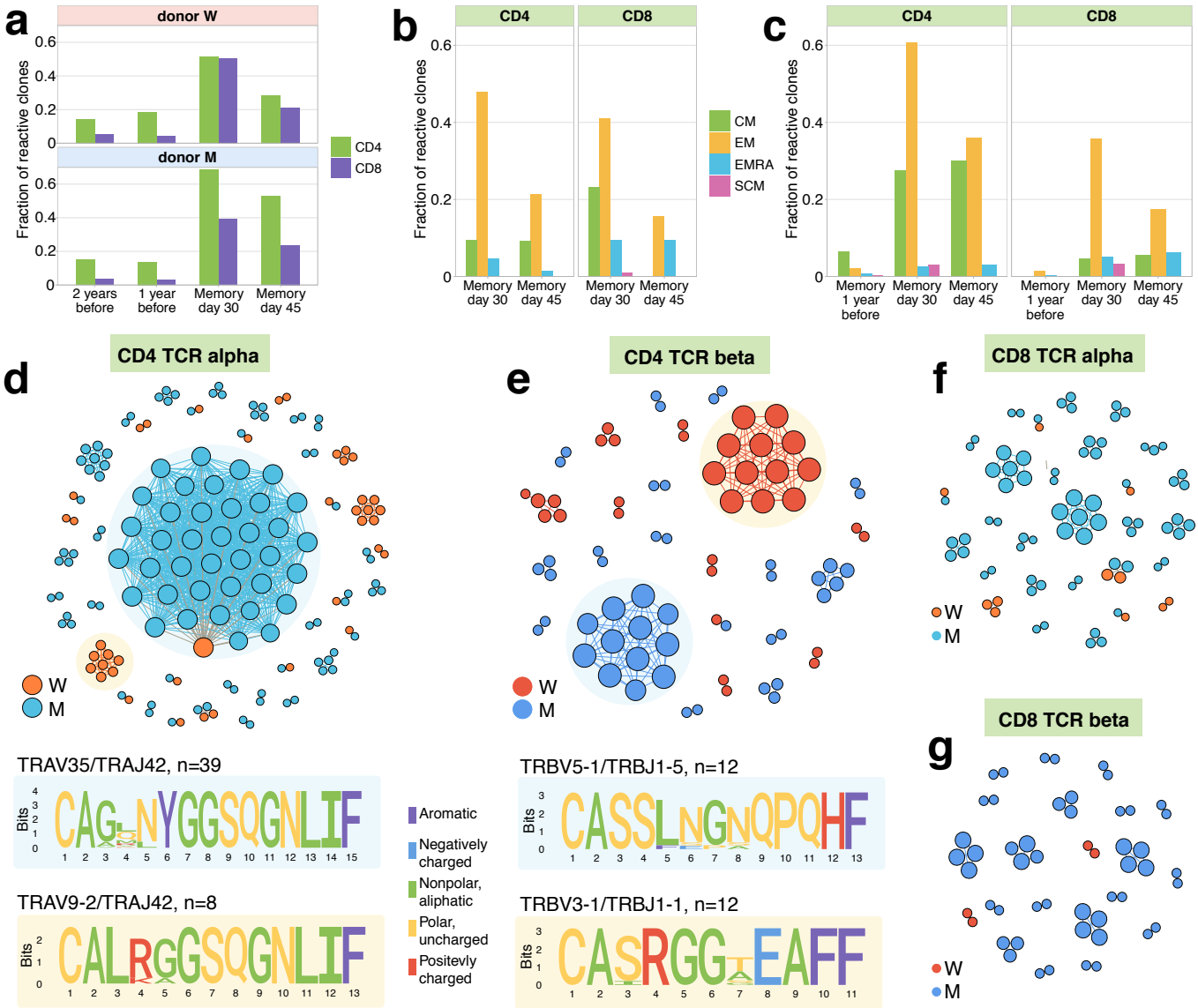

FIG. 2. **Memory phenotypes and TCR sequence motifs of responding clonotypes contracting after day 15. a, A large fraction of contracting clonotypes is identified in T cell memory subsets after infection.** Fraction of contracting CD4+ and CD8+ TCRbeta clonotypes present in 1-year and 2-year pre-infection PBMC and in at least one of memory subpopulation sampled on day 30 and day 37 post infection. **b, Responding clones are found in different memory subsets.** For both W (left) and M (right) donors, CD4+ clonotypes were more frequently found in the Central Memory (CM) subset than CD8+ T cells, and both CD4+ and CD8+ contracting clonotypes were present in the Effector Memory (EM) compartment. **c,** For donor M (right), CD4+ (but few CD8+) contracting clonotypes are also identified in memory subsets 1 year before the infection, with a bias towards the CM subpopulation. **d, Analysis of TCR amino acid sequences of contracting clones reveal distinctive motifs.** Each vertex in the similarity network corresponds to a contracting clonotype. An edge indicates 2 or less amino acid mismatches in the CDR3 region (and identical V and J segments). Networks are plotted separately for CD4alpha **(d)**, CD4beta **(e)**, CD8alpha **(f)** and CD8beta **(g)** contracting clonotypes. Clonotypes without neighbours are not shown. Sequence logos corresponding to the largest clusters are shown under the corresponding network plots.

CoV-2 or any other coronavirus-specific TCR sequences in VDJdb, so the absence of matches suggests that contracting and expanding clones are unlikely to be specific for immunodominant epitopes of common pathogens covered in VDJdb.

It was previously shown that TCRs recognising the same antigens frequently have highly similar TCR sequences [24, 25]. To identify motifs in TCR amino acid sequences, we plotted similarity networks for significantly contracted (Fig. 2d-g) and expanded (Fig.

S5d-g) clonotypes. In both donors we found clusters of highly similar clones in both CD4+ and CD8+ subsets for expanding and contracting clonotypes. Clusters were largely donor-specific, as expected, since our donors have dissimilar HLA alleles (SI Table 1) and thus each is likely to present a non-overlapping set of T cell antigens. The largest cluster, described by the motif TRAV35-CAGXNYGGSQGNLIF-TRAJ42, was identified in donor M's CD4+ contracting alpha chains. Clones from this cluster constituted 15.3% of all of donor M's CD4+ responding cells on day 15, suggesting a response to an immunodominant CD4+ epitope in the SARS-CoV-2 proteome. The high similarity of the TCR sequences of responding clones in this cluster allowed us to independently identify motifs from donor M's CD4 alpha contracting clones using the ALICE algorithm [26] (Fig. S8). While the time dependent methods (Fig. 1) identify abundant clones, the ALICE approach is complementary to both edgeR and NoisET as it identifies clusters of T cells with similar sequences independently of their individual abundances. Finding highly similar responding TCR sequences in both donors, despite their different HLA types, suggests that some of the identified responding clones could be public, and are likely to be found in other SARS-CoV-2 infected individuals. Using a mechanistic model of TCR recombination [27, 28] we predicted the probability of occurrence of the reactive clones (see Methods). For TCRbeta we also looked for the identified contracting and expanding clones in a large public dataset of TCRbeta repertoires from [29]. The TCRbeta occurrence probability correlated well with the frequency of contracting and expanding clones in this public dataset (Fig. S7), suggesting that at least a fraction of the SARS-CoV-2 responding clones are public and are likely to be shared between patients.

DISCUSSION

Using longitudinal repertoire sequencing, we identified a group of CD4+ and CD8+ T clones that contract after recovery from a coronavirus infection. Our response timelines agree with T cell dynamics reported by Theravajan et al. [9] for mild COVID-19, as well as with dynamics of T cell response to live vaccines [11]. Surprisingly, in both donors we also identified a group of predominantly CD4+ clonotypes which expanded from day 15 to day 37 after the infection. One possible explanation for this second wave of expansion is the priming of CD4+ T cells by antigen-specific B-cells, but there might be other mechanisms such as the migration of SARS-CoV-2 specific T cells from lymphoid organs. It is also possible that later expanding T cells are triggered by another infection, simultaneously and asymptomatically occurring in both donors around day 30. The major limitation of our study is that we do not know the specificity of contracted and expanded clones observed after infection. However, accumulation of TCR sequences from SARS-CoV-2-specific T cells will allow us to connect the TCR motifs we described to particular epitopes of the virus and estimate their relative contribution to the immune response. Databases of SARS-CoV-2-associated TCR sequences combined with sequencing of patients' TCR repertoires will allow for the tracking of the adaptive immune response in the early stages of the disease, and possibly identify features correlating with clinical outcomes. We showed that a large fraction of putatively SARS-CoV-2 reactive T cell clones are later found in memory subpopulations, and a subset of CD4+ clones were identified in pre-infection central memory subsets. The presence of SARS-CoV-2 cross-reactive CD4 T cells in healthy individuals was recently demonstrated [14, 15]. Our data further suggests that cross-reactive CD4+ T cells can participate in the response *in vivo*. It is interesting to ask if the presence of cross-reactive T cells before infection is linked to the mildness of the disease. Larger studies with cohorts of severe and mild cases with pre-infection timepoints are needed to address this question.

METHODS

Donors and blood samples

Peripheral blood samples from two young healthy adult volunteers, donor W (female) and donor M (male) were collected with written informed consent in a certified diagnostics laboratory. Both donors gave written informed consent to participate in the study under the declaration of Helsinki. HLA alleles of both donors (SI Table 1) were determined by an in-house cDNA high-throughput sequencing method.

SARS-CoV-2 S-RBD domain specific ELISA

An ELISA assay kit developed by the National Research Centre for Hematology was used for detection of anti-S-RBD IgG according to the manufacturer's protocol. The relative IgG level was calculated by dividing the OD (optical density) values by the mean OD value of the cut-off positive control serum supplied with the Kit. OD values of 2 samples (d37 and d45) for donor M exceeded the limit of linearity for the Kit. In order to properly compare the relative IgG levels between d30, d37 and d45, these samples were diluted 1:400 instead of 1:100, the ratios d37:d30 and d45:d30 were calculated and used to calculate the relative IgG level of d37 and d45.

Isolation of PBMCs and T cell subpopulations

PBMCs were isolated with the Ficoll-Paque density gradient centrifugation protocol. CD4+ and CD8+ T cells were isolated from PBMCs with Dynabeads CD4+

and CD8+ positive selection kits respectively. For isolation of EM, EMRA, CM and SCM memory subpopulations we stained PBMCs with the following antibody mix: anti-CD3-FITC (UCHT1, eBioscience), anti-CD45RA-eFluor450 (HI100, eBioscience), anti-CCR7-APC (3D12, eBioscience), anti-CD95-PE (DX2, eBioscience). Cell sorting was performed on FACS Aria III, all four isolated subpopulations were lysed with Trizol reagent immediately after sorting.

TCR library preparation and sequencing

TCRalpha and TCRbeta cDNA libraries preparation was performed as previously described in [19]. RNA was isolated from each sample using Trizol reagent according to the manufacturer's instructions. A universal primer binding site, sample barcode and unique molecular identifier (UMI) sequences were introduced using the 5'RACE technology with TCRalpha and TCRbeta constant segment specific primers for cDNA synthesis. cDNA libraries were amplified in two PCR steps, with introduction of the second sample barcode and Illumina TruSeq adapter sequences at the second PCR step. Libraries were sequenced using the Illumina NovaSeq platform (2x150bp read length).

TCR repertoire data analysis

Raw data preprocessing. Raw sequencing data was demultiplexed and UMI guided consensuses were built using migec v.1.2.7 [30]. Resulting UMI consensuses were aligned to V and J genomic templates of the TRA and TRB locus and assembled into clonotypes with mixcr v.2.1.11 [31]. See SI Table 2 for the number of cells, UMIs and unique clonotypes for each sample.

Identification of clonotypes with active dynamics. Principal component analysis (PCA) of clonal trajectories was performed as described before [16]. First we selected clones which were present among the top 1000 abundant in any of post-infection PBMC repertoires. Next, for each clone we calculated its frequency at each post-infection timepoint and divided this frequency by the maximum frequency of this clone for normalization. Then we performed PCA on the resulting normalized clonal trajectory matrix and identified three clusters of trajectories with hierarchical clustering with complete linkage, using Euclidean distances between trajectories. We identify statistically significant contractions and expansions with edgeR as previously described [17], using FDR adjusted $p < 0.01$ and $\log_2$ fold change threshold of 1. NoisET implements the Bayesian detection method described in [21]. Briefly, a two-step noise model accounting for cell sampling and expression noise is inferred from replicates, and a second model of expansion is learned from the two timepoints to be compared. The procedure outputs the posterior probability of expansion or contraction, and the median estimated $\log_2$ fold change, whose thresholds are set to 0.05 and 1 respectively.

TCR generative probability estimation. Models of V(D)J-recombination for the TCRalpha and TCRbeta loci were inferred for each individual using IGoR [32]. Selection models were then inferred and evaluated using SONIA [28] to estimate TCR occurrence probabilities.

Data availability

Raw sequencing data are deposited to the Short Read Archive (SRA) accession: PRJNA633317. Processed TCRalpha and TCRbeta repertoire datasets, resulting repertoires of SARS-CoV-2-reactive clones, and raw data preprocessing instructions can be accessed from: `https://github.com/pogorely/Minervina_COVID`.

ACKNOWLEDGMENTS

The study was supported by RSF 20-15-00351. E.R and A.F. are supported by the DFG Excellence Cluster Precision Medicine in Chronic Inflammation (Exc2167) and the DFG grant n. 4096610003, M.B.K., T.M., and A.M.W. are supported by European Research Council COG 724208. I.Z.M. is supported by RFBR 19-54-12011 and 18-29-09132. D.M.C. is supported by the grant from Ministry of Science and Higher Education of the Russian Federation 075-15-2019-1789.

[1] Vabret N, et al. (2020) Immunology of COVID-19: current state of the science. *Immunity* p S1074761320301837.

[2] Schmidt ME, Varga SM (2018) The CD8 T Cell Response to Respiratory Virus Infections. *Frontiers in Immunology* 9:678.

[3] Swain SL, McKinstry KK, Strutt TM (2012) Expanding roles for CD4+ T cells in immunity to viruses. *Nature Reviews Immunology* 12:136–148.

[4] Ng OW, et al. (2016) Memory T cell responses targeting the SARS coronavirus persist up to 11 years post-infection. *Vaccine* 34:2008–2014.

[5] Oh HLJ, et al. (2011) Engineering T Cells Specific for a Dominant Severe Acute Respiratory Syndrome Coronavirus CD8 T Cell Epitope. *Journal of Virology* 85:10464–10471.

[6] Zhao J, Zhao J, Perlman S (2010) T Cell Responses Are Required for Protection from Clinical Disease and for Virus Clearance in Severe Acute Respiratory Syndrome Coronavirus-Infected Mice. *Journal of Virology* 84:9318–9325.

[7] Quinti I, et al. (2020) A possible role for B cells in COVID-19? Lesson from patients with agammaglobulinemia. *Journal of Allergy and Clinical Immunology* p S0091674920305571.

[8] Soresina A, et al. (2020) Two X-linked agammaglobulinemia patients develop pneumonia as COVID-19 manifestation but recover. *Pediatric Allergy and Immunology* p pai.13263.

[9] Thevarajan I, et al. (2020) Breadth of concomitant immune responses prior to patient recovery: a case report of non-severe COVID-19. *Nature Medicine* 26:453–455.

[10] Bi Q, et al. (2020) Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *The Lancet Infectious Diseases* p S1473309920302875.

[11] Miller JD, et al. (2008) Human effector and memory CD8+ T cell responses to smallpox and yellow fever vaccines. *Immunity* 28:710–722.

[12] Ni L, et al. (2020) Detection of SARS-CoV-2-Specific Humoral and Cellular Immunity in COVID-19 Convalescent Individuals. *Immunity* p S1074761320301813.

[13] Weiskopf D, et al. (2020) Phenotype of SARS-CoV-2-specific T-cells in COVID-19 patients with acute respiratory distress syndrome. *medRxiv*.

[14] Braun J, et al. (2020) Presence of SARS-CoV-2 reactive T cells in COVID-19 patients and healthy donors. *medRxiv*.

[15] Grifoni A, et al. (2020) Targets of T cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell*.

[16] Minervina AA, et al. (2020) Primary and secondary antiviral response captured by the dynamics and phenotype of individual T cell clones. *eLife* 9:e53704.

[17] Pogorelyy MV, et al. (2018) Precise tracking of vaccine-responding T cell clones reveals convergent and personalized response in identical twins. *Proceedings of the National Academy of Sciences of the United States of America* 115:12704–12709.

[18] DeWitt WS, et al. (2015) Dynamics of the cytotoxic T cell response to a model of acute viral infection. *Journal of Virology* 89:4517–4526.

[19] Pogorelyy MV, et al. (2017) Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. *PLoS computational biology* 13:e1005572.

[20] Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* 26:139–140.

[21] Puelma Touzel M, Walczak AM, Mora T (2020) Inferring the immune response from repertoire sequencing. *PLOS Computational Biology* 16:e1007873.

[22] Fuertes Marraco SA, et al. (2015) Long-lasting stem cell-like memory CD8+ T cells with a naïve-like profile upon yellow fever vaccination. *Science Translational Medicine* 7:282ra48.

[23] Bagaev DV, et al. (2020) VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Research* 48:D1057–D1062.

[24] Dash P, et al. (2017) Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* 547:89–93.

[25] Glanville J, et al. (2017) Identifying specificity groups in the T cell receptor repertoire. *Nature* 547:94–98.

[26] Pogorelyy MV, et al. (2019) Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLOS Biology* 17:e3000314.

[27] Sethna Z, Elhanati Y, Callan CG, Walczak AM, Mora T (2019) OLGA: fast computation of generation probabilities of B- and T-cell receptor amino acid sequences and motifs. *Bioinformatics* 35:2974–2981.

[28] Sethna Z, et al. (2020) Population variability in the generation and thymic selection of T-cell repertoires. *arXiv:2001.02843 [q-bio]* arXiv: 2001.02843.

[29] Emerson RO, et al. (2017) Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nature Genetics* 49:659–665.

[30] Shugay M, et al. (2014) Towards error-free profiling of immune repertoires. *Nature Methods* 11:653–655.

[31] Bolotin DA, et al. (2015) MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods* 12:380–381.

[32] Marcou Q, Mora T, Walczak AM (2018) High-throughput immune repertoire analysis with IGoR. *Nature Communications* 9:561.

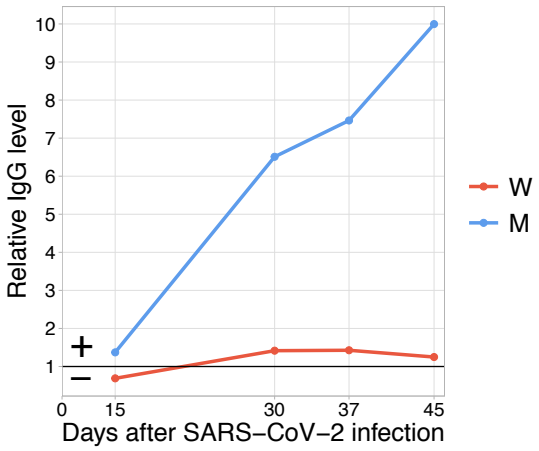

FIG. S1. **Both donors developed anti-SARS-CoV-2 IgG response by day 30 post infection.** The relative level of SARS-CoV-2 S-RBD domain specific IgG (y-axis) is plotted against time. Solid black line shows the threshold for positive testing.

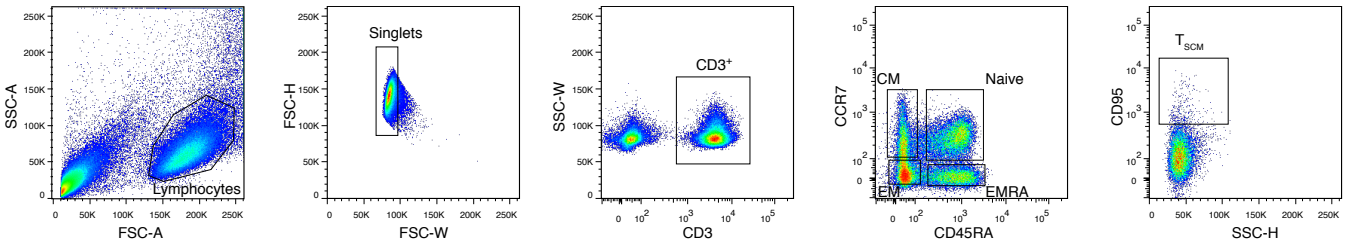

FIG. S2. **Memory subpopulation gating strategy.** Three populations of memory T cells: EM, CM and EMRA are defined by CCR7 and CD45RA markers, SCM are distinguished from naive CCR7+ CD45RA+ T cells by CD95 expression.

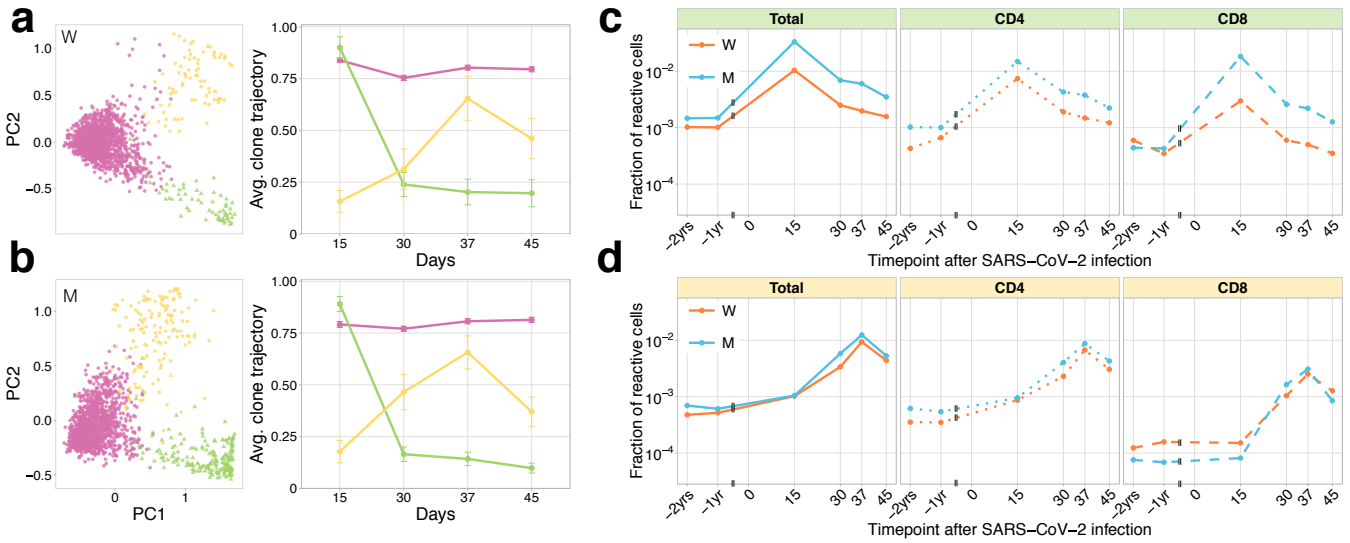

FIG. S3. **Longitudinal tracking of T cell clones after mild COVID-19 with TCRalpha repertoire sequencing. a,b, PCA of clonal temporal trajectories identifies three groups of clones with distinctive behaviours.** Left: first two principal components of 1000 most abundant TCRalpha clonotype normalized frequencies at post-infection PBMC timepoints. Right: mean +- 2.96 SE of the normalized trajectory of TCR clones in each cluster. **c,d Contracting and expanding clones include both CD4+ and CD8+ T cells and are less abundant in pre-infection repertoires.** T cell clones significantly contracted from day 15 to day 45 **(c)** and significantly expanded from day 15 to day 37 **(d)** were identified using edgeR in both donors. The fraction of contracting **(c)** and expanding **(d)** TCRalpha clonotypes in the total repertoire (corresponding to the fraction of responding cells of all T cells) is plotted in log-scale for all reactive clones (left), reactive clones with the CD4 (middle) and the CD8 phenotype (right).

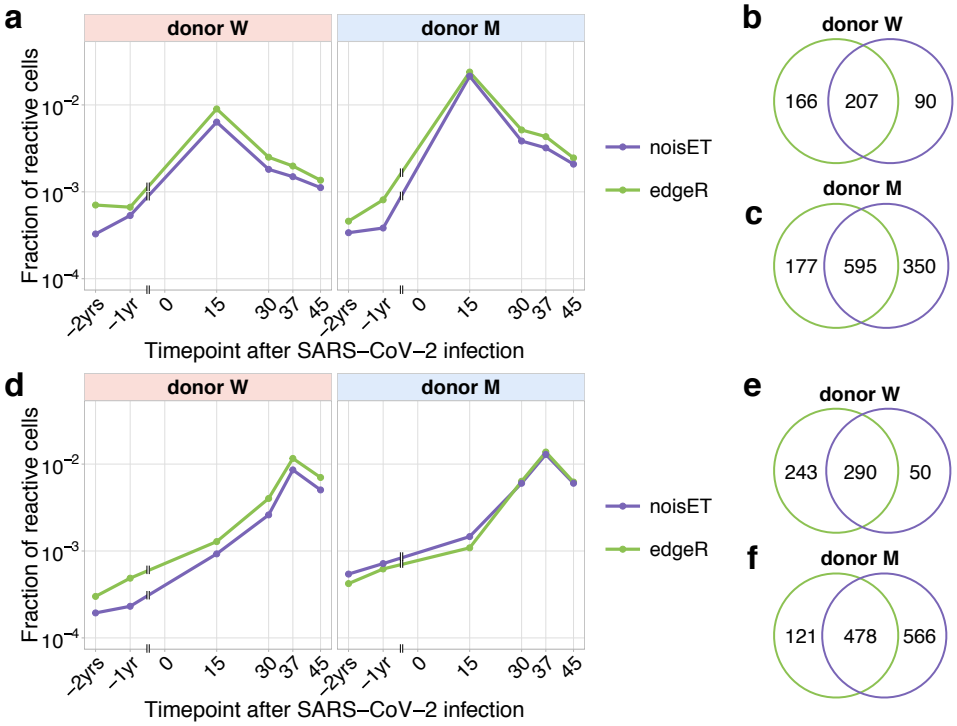

FIG. S4. **Comparison of clonal expansion detection procedures.** The fraction (plotted in the log-scale) of contracting **(a)** and expanding **(d)** TCRbeta clonotypes in the total repertoire was estimated using subsets of expanded and contracted clones called by edgeR (green) and NoisET (purple) models. Overlaps for contracted clones **(b,c)** and expanded clones **(e,f)** identified by both models are shown on the right.

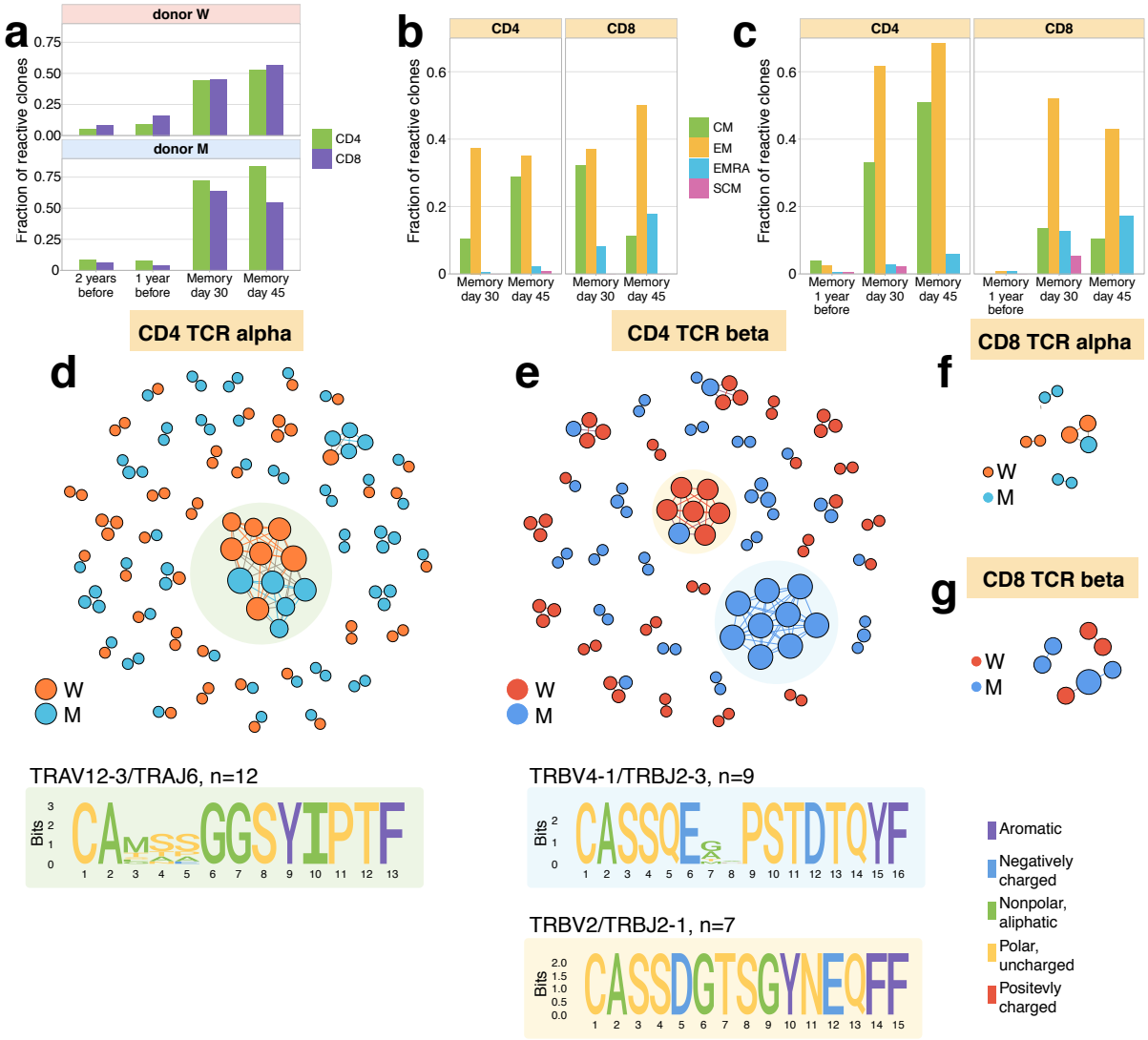

FIG. S5. **Memory phenotypes and TCR sequence motifs of responding clonotypes expanding from day 15 to day 37. a, A large fraction of expanding clonotypes is identified in T cell memory subsets after infection.** A small fraction of expanding CD4+ and CD8+ TCRbeta clonotypes present in 1-year and 2-year pre-infection PBMC and in at least one of memory subpopulation sampled on day 30 and day 37 post infection. **b, Responding clones are found in different memory subsets.** For both W (left) and M (right) donors, CD4+ clonotypes were more frequently found in the Central Memory (CM) subset than CD8+ T cells, and both CD4+ and CD8+ expanding clonotypes were present in the Effector Memory (EM) compartment. **c,** For donor M (right), CD4+ (but few CD8) expanding clonotypes are also identified in memory subsets 1 year before the infection, with a bias towards the CM subpopulation. **d, Analysis of TCR amino acid sequences of expanding clones revealed distinctive motifs.** Each vertex in the similarity network corresponds to a expanding clonotype. An edge indicates 2 or less amino acid mismatches in the CDR3 region (same V and J segments are required for connected vertices). Networks are plotted separately for CD4alpha **(d)**, CD4beta **(e)**, CD8alpha **(f)** and CD8beta **(g)** contracting clonotypes. Clonotypes without neighbours are not shown. Sequence logos corresponding to the largest clusters are shown under corresponding network plots.

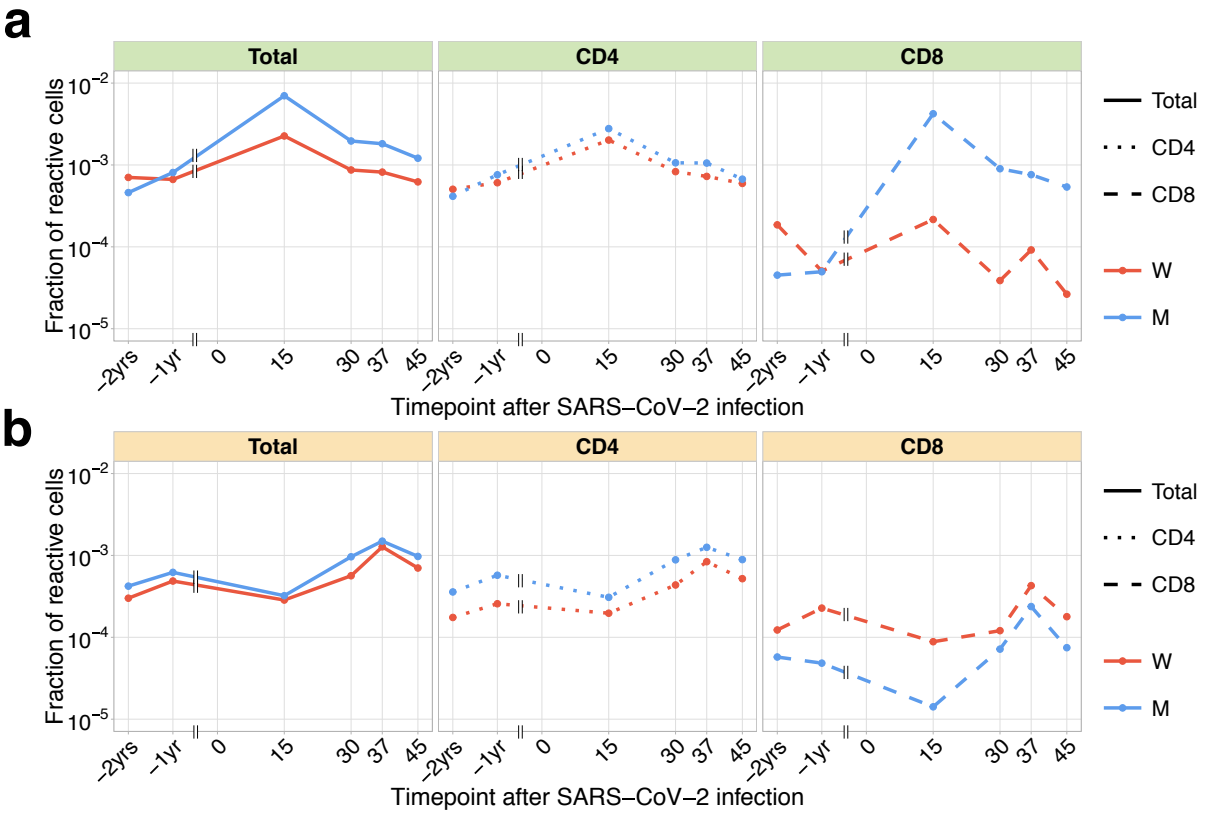

FIG. S6. **Dynamics of pre-existing SARS-CoV-2 responding clones.** The fraction of pre-existing (identified in -1 yr and/or -2 yr timepoint pre-infection) contracting **(a)** and expanding **(b)** TCRbeta clonotypes in the total repertoire (corresponding to the fraction of responding cells of all T cells) is plotted in log-scale for all reactive clones (left), reactive clones with the CD4 (middle) and the CD8 phenotype (right).

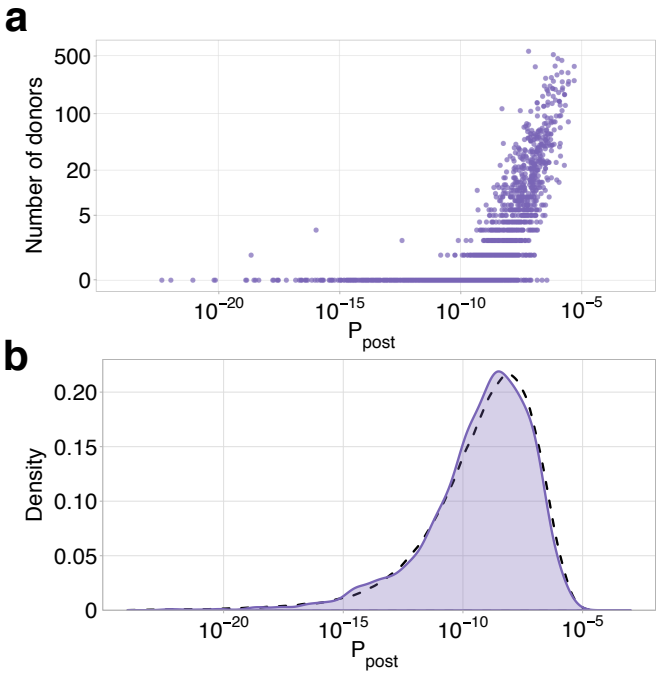

FIG. S7. **TCR occcurence probability predicts public responding clonotypes.** **a** Probability of TCR generation and selection (Ppost) of TCRbeta clones with active dynamics (both contracted and expanded from both donors) is strongly correlated with the occurrence of these clones in the large public database of TCRbeta repertoires from Emerson et al. [29]. **(b)**Distribution for TCR occurency probability of SARS-CoV-2 responding TCRbeta clonotypes. Dashed line shows TCR occurrence probability distribution for all clones in total repertoire. Some identified clones are public and have large recombination probabilities, and thus are expected to be shared with additional SARS-CoV-2 patients.

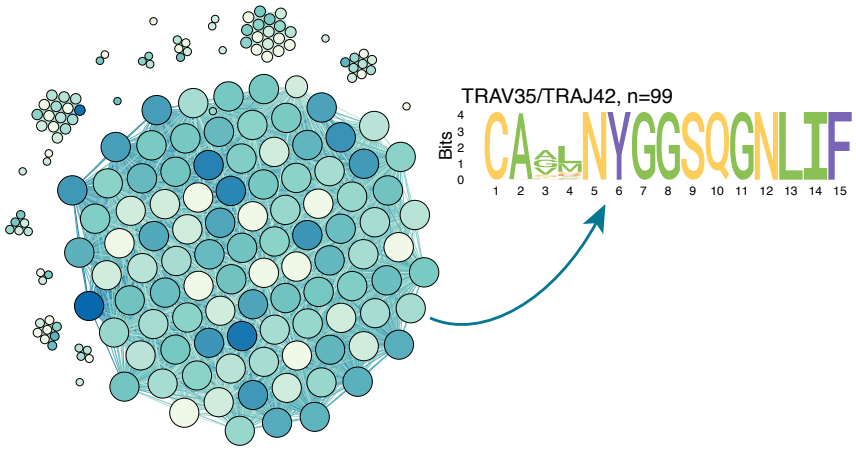

FIG. S8. **ALICE algorithm output for TCRalpha PBMC repertoire of donor M on day 15.** Similarity network shows ALICE hits (clones in repertoire with more neighbours than expected by chance), which differ by 2 mismatches or less in TCRalpha amino acid sequence. Darker colors indicate larger frequency of clone in the repertoire, vertex size indicates degree. The majority (54%, 99/183) of hits identified by the algorithm correspond to a single large TRAV35/TRAJ42 cluster of CD4+ contracting clones also seen on Fig. 1d.