Cell Reports Medicine, Volume *4*

Supplemental information

Targets and cross-reactivity of human T cell recognition of common cold coronaviruses

Alison Tarke, Yun Zhang, Nils Methot, Tara M. Narowski, Elizabeth Phillips, Simon Mallal, April Frazier, Gilberto Filaci, Daniela Weiskopf, Jennifer M. Dan, Lakshmanane Premkumar, Richard H. Scheuermann, Alessandro Sette, and Alba Grifoni

Table S1. Summary of the cohorts analyzed in this study, related to Figures 1, 3, and 4.

	First Cohort (n = 88)	Validation Cohort (n = 7)
Age (years)	19-84 [Median = 45, IQR = 30]	23-74 [Median = 59, IQR = 46]
Sex		
Male (%)	48% (43/88)	86% (6/7)
Female (%)	52% (46/88)	14% (1/7)
Sample Collection Date	March 2020 – February 2021	October 2018 – August 2019
Race-Ethnicity		
White- not Hispanic or Latino	60% (53/88)	44% (3/7)
Hispanic or Latino	15% (13/88)	14% (1/7)
Asian	6% (5/88)	14% (1/7)
Black or African American	1% (1/88)	14% (1/7)
Not reported	18% (16/88)	14% (1/7)

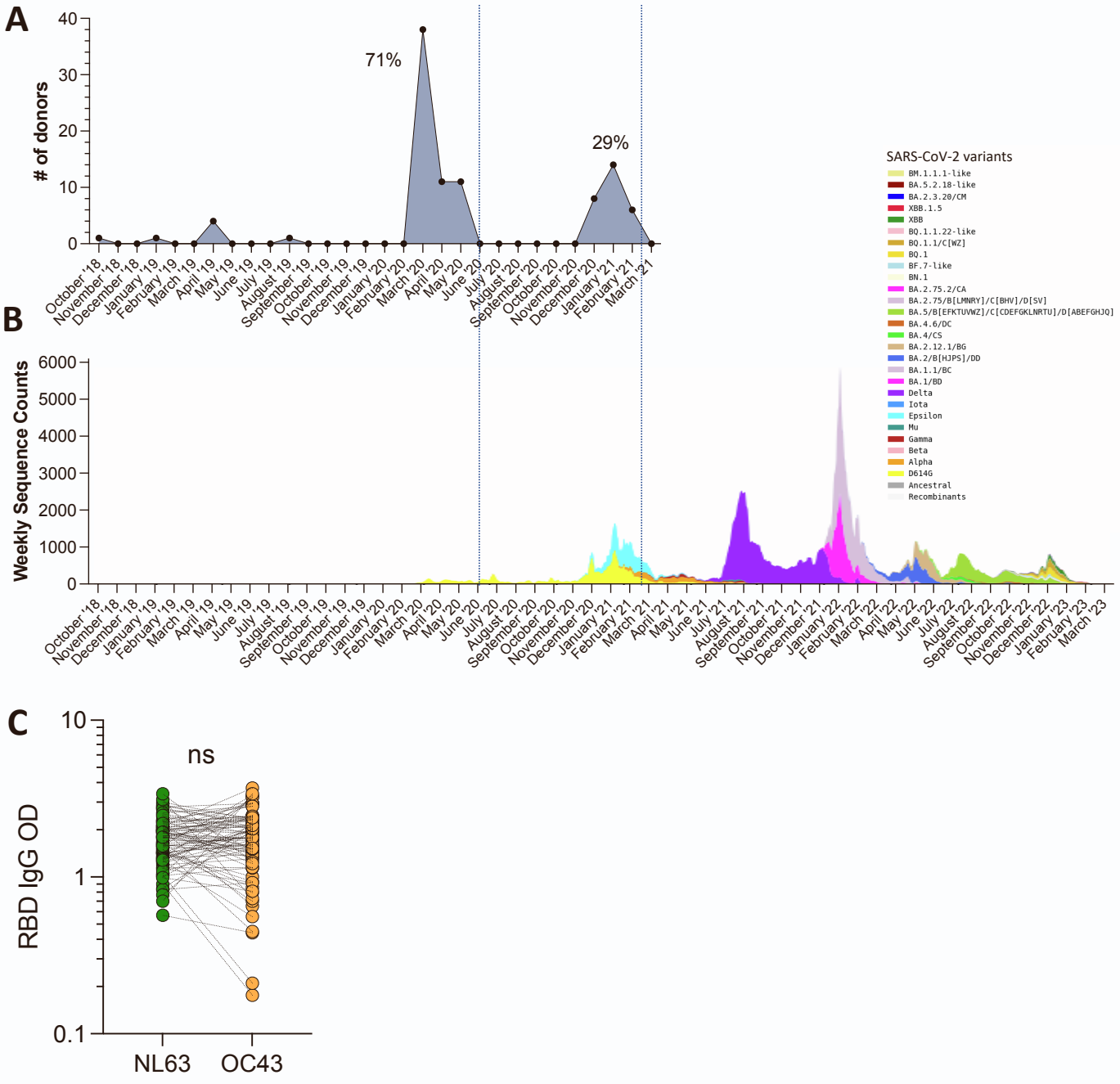


Figure S1. Collection periods and NL63 and OC43 serology of the cohort of healthy donors, related to Figures 1, 3, and 4.

Samples from a cohort of healthy donors (n=88) were collected during the period from October 2018 to Februrary 2021. **A**) 71% of samples were collected prior to June 2020 and the remaining 29% were collected from November 2020 to February 2021. **B**) Weekly sequence counts for the SARS-CoV-2 variants in San Diego during this time period were downloaded from https://cov.lanl.gov/content/sequence/EMBERS/embers.html. **C**) The RBD IgG OD ELISA titers for NL63 and OC43 are shown for the healthy donors and the dotted lines connect the same donor analyzed for NL63 or OC43. Comparison of serum antibodies to NL63 and OC43 RBD was performed using the Wilcoxon rank-sum test.

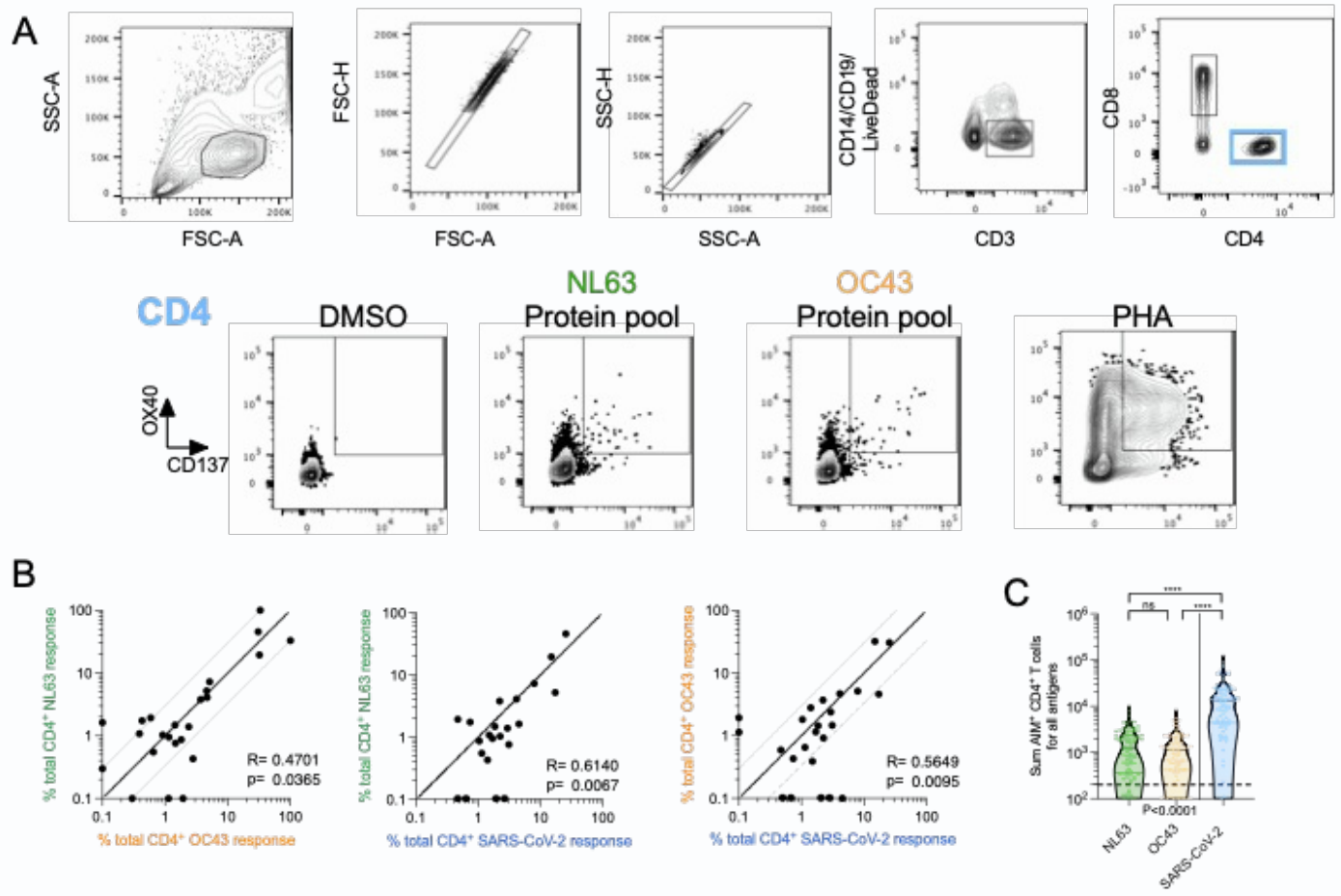
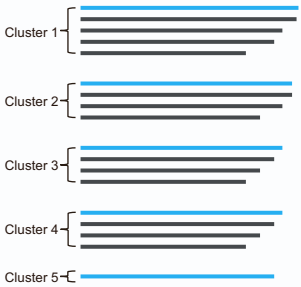


Figure S2. Total AIM$^+$ CD4$^+$ T cell reactivity against antigens related to NL63, OC43 and SARS-CoV-2, related to Table S1 and Figure 1.

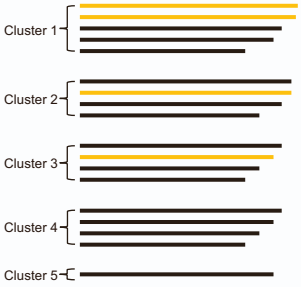
A) The gating strategy for the AIM assay is shown. **B**) Data are expressed as sum counts of OX40$^+$ CD137$^+$ CD4$^+$ T cells for each individual positive antigen for NL63 (green), OC43 (orange) and SARS-CoV-2 (blue). Historical data on T cell responses to SARS-CoV-2 is from COVID-19 convalescent donors (n=99) originally published in Tarke et al. (Tarke et al., 2021a). OC43 and NL63 T cell reactivity reflects the First Cohort of healthy donors (n=88) described in this study. Pairwise correlation among the three viruses per protein is shown together with Spearman correlation R and p-value. **C**) Comparison of the three viruses are shown. Data are compared by Kruskal Wallis test (P<0.0001, below) as well as Mann-Whitney (above) for each of the paired comparisons. **** $p < 0.0001$.



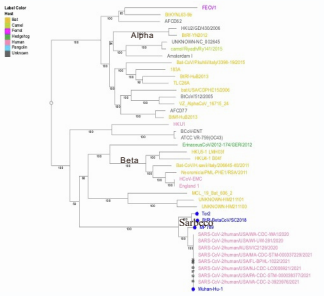
Sequence identity-based clustering using cd-hit's non-greedy option.
The seed sequence (blue) in each cluster was selected as the representative by cd-hit.

Metadata

Taxa and metadata associated with the sequences were extracted and subsequently used for re-selecting representative viruses.



Target clusters (clusters 1-3) were selected based on cluster size and host. Next, representatives (orange) were selected for each target cluster considering sequence and annotation quality, host, isolation date and region, and RefSeq designation.



For taxon of finer granularity such as SARS-CoV-1/2, sequences are highly similar. To ensure the best accuracy and diversity representation, a phylogeny-based sampling method was used to select representatives (blue) in major phylogenetic clusters.

Figure S3. Representative virus selection, related to Table 1.
To select representative viruses for each taxon group (alphaCoV, sarbeco, non-sarbeco betaCoV), a targeted sampling approach was used which leverages sequence identity, sequence and annotation quality, host, isolation date and region, RefSeq designation, and phylogenetic structure (see Methods for details).

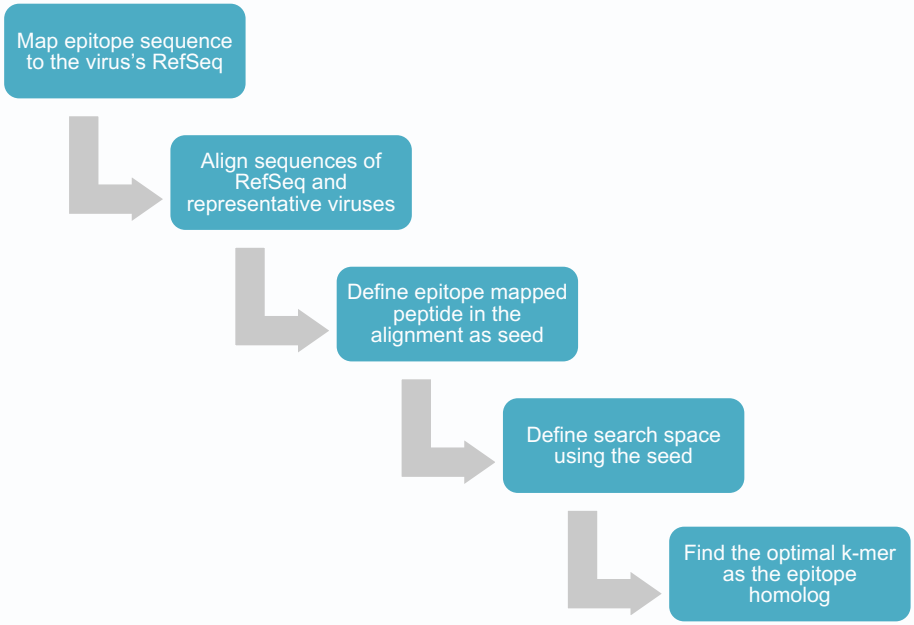


Figure S4. **Pipeline to establish the degree of sequence conservation of each epitope, related to Figure 2**.
Epitopes were mapped to each of the representative viruses to identify the epitope homologs using an alignment-based, k-mer finding approach (see Methods for details). This process first identified the epitope homologous region in each representative and then selected the optimal k-mer in the region as the epitope homolog. The level of sequence conservation of the homologous epitope regions is shown in **Figure 2**.