

Figure 1: Architecture of the SARS-CoV-2 genome and proteome. Virion structural proteins spike (S), envelope (E), membrane (M) and nucleocapsid (N) are showed in purple. Current vaccines only elicit immunity to spike protein. Non-structural proteins are in shades of blue and open reading frame (Orf) proteins appear in green. Figure taken from ?

In an attempt to shed light in TCR biomarker discovery, researchers have used *in vitro* antigen-enrichment of TCR repertoires, such as peptide-MHC tetramer sorting assays or Multiplexed Identification of T cell Receptor Antigen specificity assays (MIRA). In brief, antigen-specific TCR discovery with this techniques consists in exposing T cells from a blood sample to known epitopes fixed in a reagent and then sequence the TCRs of the responding T cells (?). Then, this antigen-specific TCRs can be used to screen bulk TCR repertoires in a cohort of interest, but the constraints of high inter-individual diversity and TCR cross-reactivity may still make it difficult to find these TCRs in the population (?).

For many pathologies the major impediment to their study from an immunological repertoires perspective is the lack of these *in-vitro*-determined antigen-specific TCRs (?), but the scenario is somewhat better for COVID-19. In July 2020 Adaptive Biotechnologies and Microsoft publicly released data of MIRA assays including more than 160000 SARS-CoV-2-specific TCRs recognizing 269 different SARS-CoV-2 epitopes (?). These epitopes covered most of SARS-CoV-2 proteins, shown in Fig. 1.

Several articles about SARS-CoV-2 vaccines eliciting robust antibody responses against SARS-CoV-2 spike (S) protein were published in high impact journals during early phase clinical trials (???) and later came the first preprints analyzing the immunogenicity of vaccines from a T cell and TCR repertoire perspective (???). However, these studies tend to analyze TCR repertoires in a very narrow manner, limiting their analysis to measure the frequencies of the SARS-CoV-2-specific TCRs found in a repertoire by an exact sequence match, and therefore probably underestimating the TCR response in COVID-19.

The present study aims to analyze in depth the published TCR repertoires of 19 COVID-19 convalescent individuals, 19 Ad26.COV2.S vaccine (developed by Janssen Pharmaceutica) recipient individuals and 5 placebo recipient individuals to fully characterize how viral and vaccine immunization alter the complex landscape of TCR repertoires in health and disease.

Material and Methods

TCR data

This study is based on public data from three previous works (???).

The dataset used for TCR repertoire analysis (?) includes samples from 32 individuals: 8 convalescent from COVID-19, 19 who received the Ad26.COV2.S vaccine developed by Janssen Pharmaceutica during a clinical trial, and 5 subjects who received a placebo. Peripheral blood samples were collected post diagnosis or vaccination and immunosequencing of the CDR3 regions of human TCR β chains was performed with the immunoSEQ Assay (Adaptive Biotechnologies). Data was accessed on July 2021

via Adaptive Biotechnologies immuneACCESS® database (immuneACCESS® DOI: <https://doi.org/10.21417/GA2021N>).

To match the sample size of vaccinated individuals with data generated with the same procedure, 11 TCR repertoire samples from COVID-19-convalescent subjects were randomly selected from the COVID-19-HUniv12Oct dataset on Adaptive Biotechnologies ImmuneCODE™ project (?). The full dataset contains TCR β repertoires from 193 convalescent patients whose blood sample was collected at the Hospital Universitario 12 de Octubre (Madrid, Spain). Data was accessed on Aug 2021 via Adaptive Biotechnologies immuneACCESS® database (immuneACCESS® DOI: <https://doi.org/10.21417/ADPT2020COVID>, ImmuneCODE-COVID-Release-002).

SARS-CoV-2-specific CD8 $^{+}$ TCR β sequences were obtained from ?. This sequences are proven to bind SARS-CoV-2 epitopes by MIRA assays (?) and are also enriched in bulk TCR β repertoires of convalescent individuals compared to healthy controls. For the present study, only TCR β sequences with a strong evidence of HLA restriction (N = 1831) were taken into consideration.

SARS-CoV-2-specific CD4 $^{+}$ TCRs discovery

While SARS-CoV-2-specific CD4 $^{+}$ have been used to annotate TCR repertoires in previous studies (??), those enriched and high-reliable datasets are not currently public. ImmuneCODE™ project contains an unenriched dataset of 6809 CD4 $^{+}$ TCRs that bind 49 different SARS-CoV-2 epitopes presented by class II MHC molecules in MIRA assays. Data was accessed on Aug 2021 via Adaptive Biotechnologies immuneACCESS® database (immuneACCESS® DOI: <https://doi.org/10.21417/ADPT2020COVID>, ImmuneCODE-COVID-Release-002).

These TCRs were further screened for enrichment compared to a background of healthy individuals repertoires in order to remove TCRs that may be highly public or cross-reactive to common antigens. 64 TCRs were selected to annotate the repertoires, in addition to CD8 $^{+}$ dataset. The enrichment analysis was performed with tcrdist3 Python toolkit (Docker image v0.1.9) (??), following the same meta-clonotype discovery pipeline employed for SARS-CoV-2 CD8 $^{+}$ TCR discovery as in ?.

Measurement of T-cell response to SARS-CoV-2

The 43 TCR β repertoires were annotated for antigen-specificity with the SARS-CoV-2-specific TCRs (CD4 $^{+}$ and CD8 $^{+}$) by exact matching of CDR3 aminoacid sequence and V gene. The SARS-CoV-2 response of each individual to spike and non-spike proteins was measured in terms of breadth, defined as the proportion of distinct TCRs recognizing certain protein among all the unique sequences in a repertoire, and in terms of depth, which is the proportion of the frequency of those SARS-CoV-2-specific TCRs.

CDR3 pairwise distances

Pairwise distances between all CDR3 sequences in a given sample were computed with tcrdist3 (??), which implements a custom distance metric based on BLOSUM62 substitution matrix to account for similar aminoacid substitutions, and applies different weights depending on the importance of every CDR3 position in antigen binding. Total runtime was \square 103 hours with parallel processing (40 CPUs, 256 GB of RAM).

Network analysis

In this analysis each unique CDR3 aminoacid sequence were considered as a node. An edge was built between two nodes if their pairwise distance was \leq 12. The reason behind this threshold is that 12 is the greatest possible distance between two CDR3 with one mismatch according to tcrdist3 algorithm. Networks were built and analyzed with R igraph package v1.2.6 (?).

Data analysis and visualization

All plots and analyses were carried in R 3.6.1 (?). For data analysis, the packages dplyr v1.0.2 (?), tidyverse v1.1.2 (?), rstatix v0.7.0 (?) and parallel v3.6.1 (?) were used. Networks were plotted with the ggraph v2.0.2 package (?). All other plots were generated with ggplot2 v3.3.2 (?) and ggpunr v0.4.0 (?). 3D

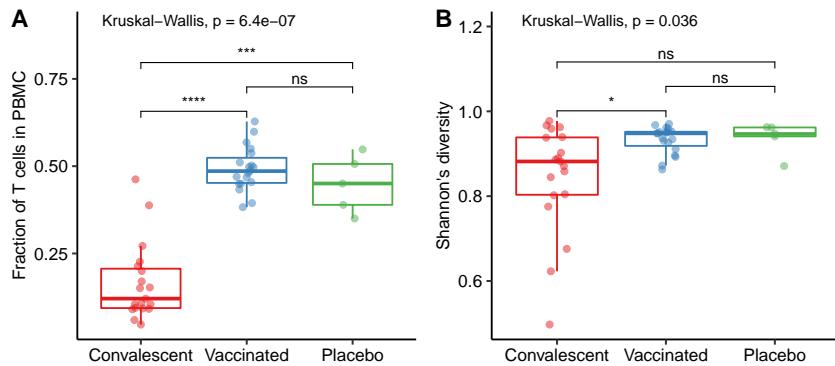


Figure 2: TCR repertoire preliminary analysis. **(a)** Fraction of T cells among peripheral blood mononuclear cells (PBMC). **(b)** Shannon diversity index. A value closer to 0 indicates the emergence of a few dominant clones, whereas it reaches its maximum when TCR frequencies are evenly distributed. Statistical significance was determined by two-sided Wilcoxon rank-sum tests. N = 43 independent samples (19 SARS-CoV-2 convalescent individuals, 19 Ad26.COV2.S vaccine recipients, 5 placebo recipients).

visualization of the SARS-CoV-2 spike protein (PDB ID: 6XR8) was generated with Protein Imager (?).

Results

TCR repertoire preliminary analysis

One of the clinical characteristics of SARS-CoV-2-infected patients, lymphopenia, can be observed from a TCR repertoire analysis perspective. In TCR repertoire sequencing from a peripheral blood sample, both the number of nucleated cells and total T cells can be estimated by the amplification of reference gene primers. The fraction of T cells is significantly lower in convalescent individuals compared to vaccinated and placebo ($p = 2.5 \cdot 10^{-9}$ and $p = 5.6 \cdot 10^{-4}$, two-sided Wilcoxon rank-sum test) and no significant differences were observed between vaccinated and placebo subjects (Fig. 2a). Some convalescent patients TCR β repertoires have a low Shannon diversity index (Fig. 2b), indicating that a few clones are expanded and possibly reflecting that these individuals had a recent adaptive immunity response, most likely to SARS-CoV-2 infection.

Breadth and depth of SARS-CoV-2-specific T-cell response

To evaluate the magnitude of the T-cell response to SARS-CoV-2 after disease and vaccination, TCR repertoires of convalescent, vaccinated and placebo recipients individuals were annotated with CD8 $^+$ and CD4 $^+$ TCR datasets that had previously been determined to be SARS-CoV-2-specific and screened for enrichment compared to a background of healthy individuals repertoires in order to remove TCRs that may be unspecific (i.e. cross-reactive to common antigens) (See Material and Methods). Among the 43 repertoires analyzed there were 11,604,850 unique TCR sequences (V gene + CDR3 aminoacid sequence) of which only 284 were SARS-CoV-2-specific. Out of these annotated 284 TCRs, 51 were public (i.e. present in more than one individual).

Response to SARS-CoV-2 spike and non-spike was measured in terms of breadth (unique TCR sequences) and depth (frequency of those TCRs). Both convalescent and vaccinated subjects have a higher spike-specific response compared to placebos in terms of breadth and depth (Fig. ??a). By contrast, breadth and depth of non-spike TCRs were significantly higher in convalescent individuals versus vaccinated and placebos, and there were no significant differences between the latter two (Fig. ??b), as expected because Ad26.COV2.S vaccine only carry the spike antigen.

An in-depth analysis of the genomic localization of SARS-CoV-2 epitopes recognized by TCRs revealed that most of the T-cell immune response in convalescent individuals is directed towards structural proteins S and N (Fig. ??a). Although vaccinated subjects show some response to non-spike proteins, for most epitopes this signal is residual compared to the convalescent group and most likely due to unspecific annotations, since placebo recipients also show a minimal level of response. The localization of the 5 spike protein epitopes is shown in Fig. ??b. As opposed to B-cell epitopes, T-cell epitopes localization is not restricted to the protein surface since TCRs recognize antigens processed and presented in MHC molecules by human cells, and it shows in the spike protein 3D representation, where most of the epitopes are partially (S1, S3, S5) or completely (S2, S4) buried in the structure. S2 epitope is in fact the most widely recognized by convalescent (9/19) and vaccinated (12/19) individuals, and S4, the most hidden

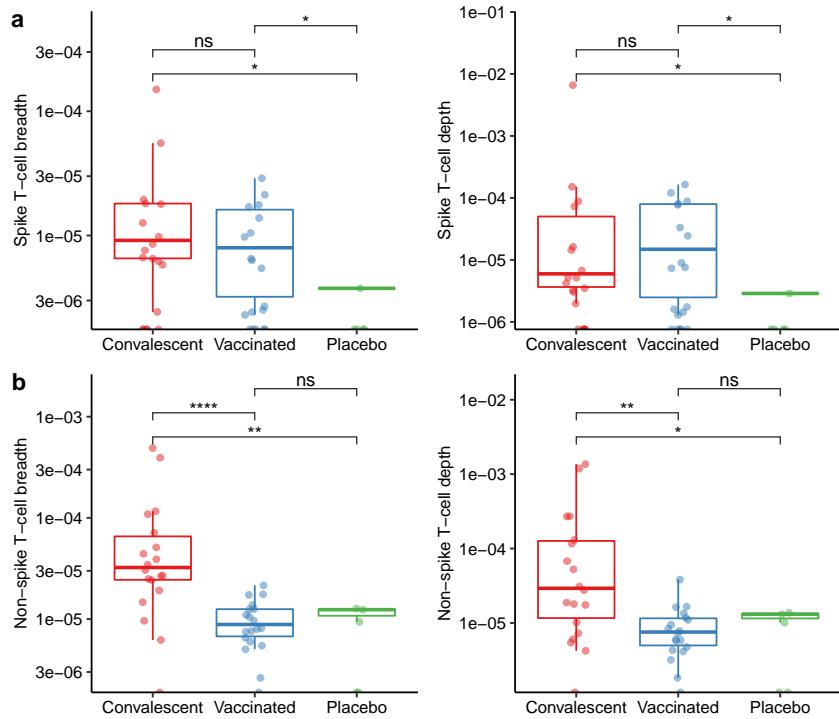


Figure 3: SARS-CoV-2-specific TCR β repertoire analysis. (a) Spike-specific T-cell breadth and depth. (b) Non-spike-specific T-cell breadth and depth. Breadth is calculated as the fraction of unique TCR sequences specific to spike / non-spike proteins; depth is the relative frequency of those specific TCRs in the repertoire. Statistical significance was determined by two-sided Wilcoxon rank-sum tests. N = 43 independent samples (19 SARS-CoV-2 convalescent individuals, 19 Ad26.COV2.S vaccine recipients, 5 placebo recipients)

in the protein, is unrecognized by convalescent subjects and exclusively recognized by TCRs of 2 out of 19 vaccine recipients.

Network analysis

The landscape of TCR repertoires is vast and complex. A simple antigen specificity annotation by V gene and CDR3 aminoacid sequence exact match, although informative and straightforward, can underestimate the magnitude of the T cell response. In order to capture the SARS-CoV-2-specific TCR repertoire architecture, graphs representing networks of similar TCRs were generated for the 43 TCR repertoires. Fig. ??a shows two SARS-CoV-2-specific TCR sequence similarity networks of one convalescent and one vaccinated individual. A SARS-CoV-2-specific TCR similarity network is defined as all the components (subgraphs in which any two nodes are connected to each other by paths) that include at least one SARS-CoV-2-specific node (yellow, purple). Spike-specific nodes are present in both graphs, while non-spike-specific nodes are restricted to the convalescent individual, as expected. In both networks, SARS-CoV-2-specific nodes have a high relative frequency. As for the networks architecture, the most striking difference is the network size, with the convalescent graph having less edges and nodes than the vaccinated one. This is also observed at the group level, and furthermore, convalescent SARS-CoV-2-specific TCRs networks have a lower average degree (number of edges per node) compared to vaccine and placebo recipient individuals (Fig. ??b).

In this work, it has been noted that direct annotation with SARS-CoV-2-specific TCRs is not very precise, since there is some non-spike response in vaccine recipients, as well as response to some SARS-CoV-2 epitopes in healthy placebo recipients (Figs. ??b, ??a). Network analysis can help identifying those nodes that truly are SARS-CoV-2-specific versus those that may be cross-reactive. It is expected for spike-specific TCRs to have more importance in convalescent and vaccinated subjects networks compared to placebo, and for non-spike-specific TCRs to be relevant only in convalescent individuals. Important nodes in network science are known as hubs, or nodes with a number of edges exceeding the average, and the network local property that measures this quality is the hub score, which takes into account how well connected is a node itself, and those nodes that it is connected to. In convalescent and placebo subjects, there were no significant differences in hub scores distributions of spike-specific and non-spike specific nodes, being the hub scores very low in the placebo recipient SARS-CoV-2-specific nodes. However, hub scores were significantly greater for spike-specific nodes in vaccinated individuals compared to non-spike-specific nodes ($p = 2.1 \cdot 10^{-3}$, two-sided Wilcoxon rank-sum test) (Fig. ??c).

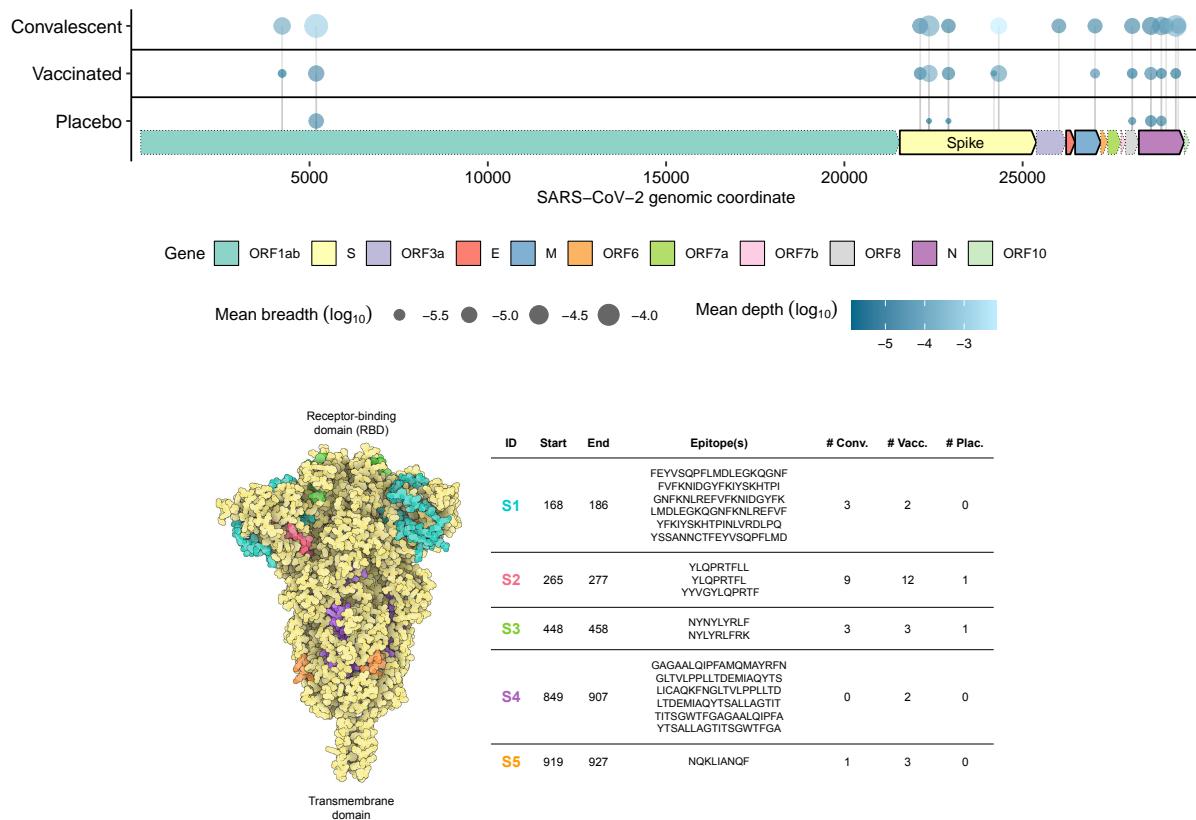


Figure 4: Epitopes recognized by T cells across SARS-CoV-2 genome. (a) Lollipop plot of the TCR β SARS-CoV-2 specificity in convalescent, vaccinated and placebo individuals across the coronavirus genome. Size of the dots indicate the mean breadth across all samples in a group and color scale indicates the mean depth of the response. Genes outlined with solid lines are structural (S, E, M and N), whereas dotted ones encode non-structural proteins. (b) Localization and characteristics of the 5 SARS-CoV-2 spike epitopes recognized by TCRs in this study. Epitopes appear colored in the three monomers of a spike protein 3D representation (PDB: 6XR8, side view). Start, End: epitope protein coordinates (1-based); # Conv., # Vacc., # Plac.: number of individuals in a group with TCRs specific to that epitope.

This suggests that TCRs annotated as non-spike-specific in vaccine recipient subjects or any SARS-CoV-2-specific TCR in the placebo group do not have an important role in their TCR repertoires and may be SARS-CoV-2-unspecific and cross-reactive.

In addition, the spike-specific nodes with highest degrees (more connections) are from vaccinated subjects, whereas the highly connected non-spike-specific nodes belong to convalescent individuals (Fig. ??d, top panels). The same trend is observed for self-loops, which represent TCR convergence, i.e. additional nucleotide TCR sequences encoding the same V gene + amino acid sequence (Fig. ??d, bottom panels). Together, all these network properties facilitate the differentiation between putative SARS-CoV-2-specific TCRs (spike-specific detected in convalescent and vaccinated individuals and non-spike-specific detected in convalescent subjects) and putative SARS-CoV-2-unspecific TCRs (those detected in placebo repertoires and non-spike-specific in vaccine recipients) (Fig. ??e). Nevertheless, due to the constantly changing nature of the TCR repertoires, after a time period following antigen exposure, some putative SARS-CoV-2-specific TCRs can be undistinguishable from putative unspecific ones in terms of hub score, degree and self-loops, since they become less important in the adaptive immunity landscape once the infection is cleared.

Bibliography

Galit Alter, Jingyou Yu, Jinyan Liu, Abishek Chandrashekhar, Erica N Borducchi, Lisa H Tostanoski, Katherine McMahan, Catherine Jacob-Dolan, David R Martinez, Aiquan Chang, et al. Immunogenicity of Ad26. COV2. S vaccine against SARS-CoV-2 variants in humans. *Nature*, pages 1–5, 2021.

Gabor Csardi and Tamas Nepusz. The igraph software package for complex network research. *InterJournal, Complex Systems*: 1695, 2006. URL <https://igraph.org>.

Pradyot Dash, Andrew J Fiore-Gartland, Tomer Hertz, George C Wang, Shalini Sharma, Aisha Souquette, Jeremy Chase

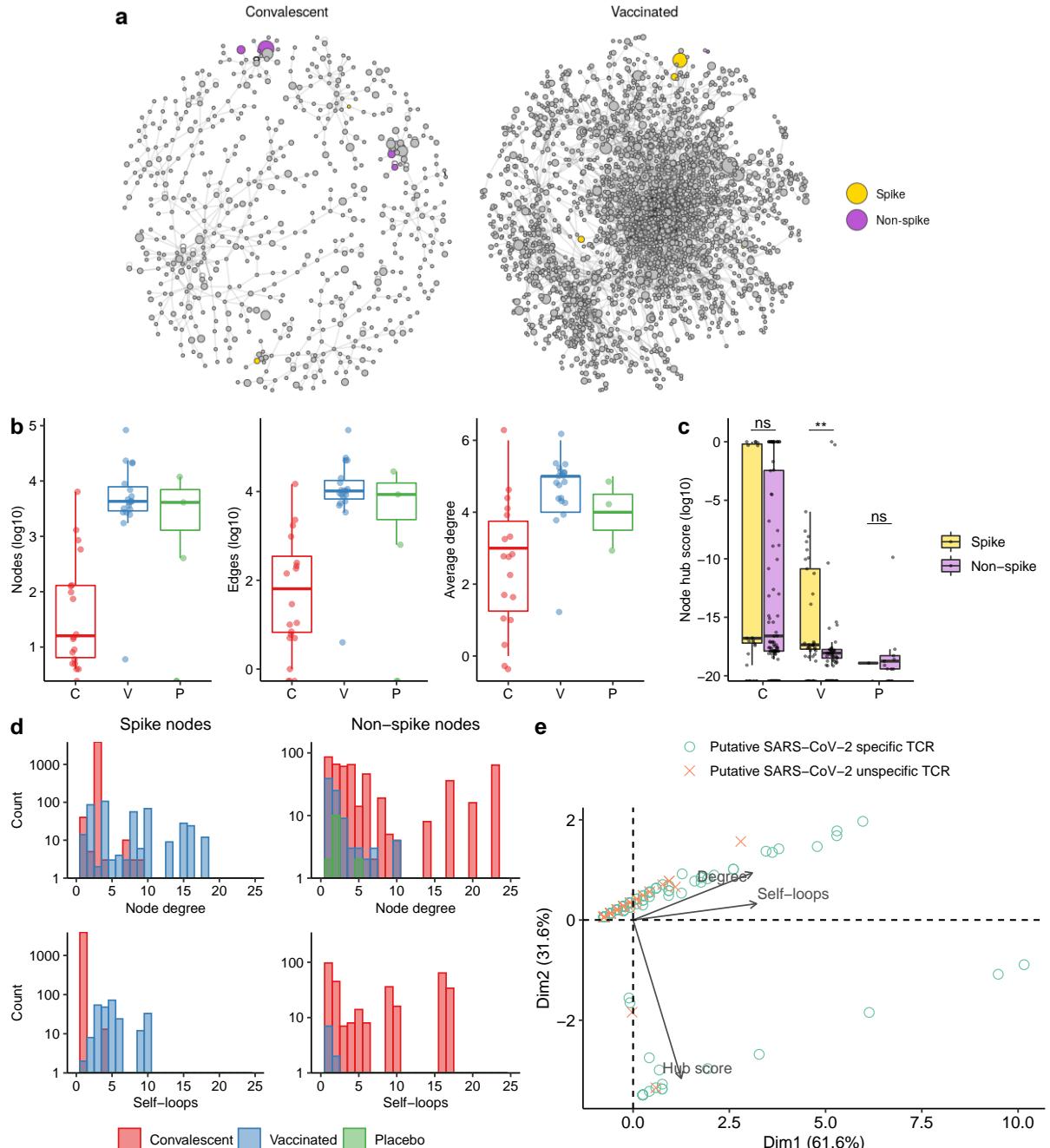


Figure 5: SARS-CoV-2-specific TCR β similarity networks. (a) TCR β similarity networks as undirected graphs, where each node represents a unique V gene + CDR3 aminoacid sequence combination. Only SARS-CoV-2-specific nodes (colored) and the nodes in their components (gray) are shown for convalescent subject 18 and vaccinated subject 12. Two nodes are connected by an edge if their tcrdist3 distance is ≤ 12 . Loops (self-edges that start and end at the same node) represent additional unique nucleotide sequences encoding the same V gene + amino acid sequence (i.e., convergence). Node size represents the TCR frequency. (b) Distribution of network global properties by group. (c) Spike and non-spike nodes hub score distributions by group. (d) Histograms of SARS-CoV-2-specific nodes degree and self-loops. Y axis represent node counts taking their frequency into account. (e) Putative SARS-CoV-2-specific and unspecific biplot based on network local metrics. A node was putatively considered SARS-CoV-2-specific if it was a spike node in a convalescent or vaccinated network, or if it was a non-spike node in a convalescent network. C: convalescent, V: vaccinated, P: placebo.

Crawford, E Bridie Clemens, Thi HO Nguyen, Katherine Kedzierska, et al. Quantifiable predictive features define epitope-specific t cell receptor repertoires. *Nature*, 547(7661):89–93, 2017.

Katie J Ewer, Jordan R Barrett, Sandra Belij-Rammerstorfer, Hannah Sharpe, Rebecca Makinson, Richard Morter, Amy Flaxman, Daniel Wright, Duncan Bellamy, Mustapha Bittaye, et al. T cell and antibody responses induced by a single dose of chadox1 ncov-19 (azd1222) vaccine in a phase 1/2 clinical trial. *Nature medicine*, 27(2):270–278, 2021.

K Christopher Garcia and Erin J Adams. How the t cell receptor sees antigen—a structural view. *Cell*, 122(3):333–336, 2005.

Rachel M Gittelman, Enrico Lavezzo, Thomas M Snyder, H Jabran Zahid, Rebecca Elyanow, Sudeb Dalai, Ilan Kirsch, Lance

Baldo, Laura Manuto, Elisa Franchin, et al. Diagnosis and tracking of sars-cov-2 infection by t-cell receptor sequencing. *medRxiv*, pages 2020–11, 2021.

Victor Greiff, Gur Yaari, and Lindsay Cowell. Mining adaptive immune receptor repertoires for biological and clinical information using machine learning. *Current Opinion in Systems Biology*, 2020.

Tim Harris and Karsten Sauer. Are T cell repertoires useful as diagnostics for SARS-CoV-2 infection? *Expert Review of Molecular Diagnostics*, 21(2):137–139, 2021.

James M Heather, Mazlina Ismail, Theres Oakes, and Benny Chain. High-throughput sequencing of the t-cell receptor repertoire: pitfalls and opportunities. *Briefings in bioinformatics*, 19(4):554–565, 2018.

Ben Hu, Hua Guo, Peng Zhou, and Zheng-Li Shi. Characteristics of sars-cov-2 and covid-19. *Nature Reviews Microbiology*, 19(3):141–154, 2021.

Alboukadel Kassambara. *ggnpubr: 'ggplot2' Based Publication Ready Plots*, 2020. URL <https://CRAN.R-project.org/package=ggnpubr>. R package version 0.4.0.

Alboukadel Kassambara. *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*, 2021. URL <https://CRAN.R-project.org/package=rstatix>. R package version 0.7.0.

Mark Klinger, Francois Pepin, Jen Wilkins, Thomas Asbury, Tobias Wittkop, Jianbiao Zheng, Martin Moorhead, and Malek Faham. Multiplex identification of antigen-specific t cell receptors using a combination of immune assays and immune receptor sequencing. *PLoS One*, 10(10):e0141561, 2015.

Jing Liu, Sumeng Li, Jia Liu, Boyun Liang, Xiaobei Wang, Hua Wang, Wei Li, Qiaoxia Tong, Jianhua Yi, Lei Zhao, et al. Longitudinal characteristics of lymphocyte responses and cytokine profiles in the peripheral blood of SARS-CoV-2 infected patients. *EBioMedicine*, 55:102763, 2020.

Joseph H Lubin, Christine Zardecki, Elliott M Dolan, Changpeng Lu, Zhuofan Shen, Shuchismita Dutta, John D Westbrook, Brian P Hudson, David S Goodsell, Jonathan K Williams, et al. Evolution of the SARS-CoV-2 proteome in three dimensions (3D) during the first six months of the COVID-19 pandemic. *bioRxiv*, 2020.

Grant Lythe, Robin E Callard, Rollo L Hoare, and Carmen Molina-París. How many tcr clonotypes does a body maintain? *Journal of theoretical biology*, 389:214–224, 2016.

Koshlan Mayer-Blackwell, Stefan Schattgen, Liel Cohen-Lavi, Jeremy Chase Crawford, Aisha Souquette, Jessica A Gaevert, Tomer Hertz, Paul G Thomas, Philip Bradley, and Andrew Fiore-Gartland. TCR meta-clonotypes for biomarker discovery with tcrdist3: quantification of public, HLA-restricted TCR biomarkers of SARS-CoV-2 infection. *bioRxiv*, 2020.

Anastasia A Minervina, Mikhail V Pogorelyy, Allison M Kirk, Emma Kaitlynn Allen, Kim J Allison, Chung-Yang Lin, David C Brice, Xun Zhu, Kasi Vegesana, Gang Wu, et al. Convergent epitope-specific t cell responses after sars-cov-2 infection and vaccination. *medRxiv*, 2021.

K.M. Murphy and C. Weaver. *Janeway's Immunobiology*. Garland Science, 2016. ISBN 9780815345503.

Sean Nolan, Marissa Vignali, Mark Klinger, Jennifer N Dines, Ian M Kaplan, Emily Svejnoha, Tracy Craft, Katie Boland, Mitch Pesesky, Rachel M Gittelman, et al. A large-scale database of T-cell receptor beta (TCR β) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. *Research square*, 2020.

Thomas Lin Pedersen. *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*, 2020. URL <https://CRAN.R-project.org/package=ggraph>. R package version 2.0.2.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL <https://www.R-project.org/>.

Ugur Sahin, Alexander Muik, Evelyn Derhovanessian, Isabel Vogler, Lena M Kranz, Mathias Vormehr, Alina Baum, Kristen Pascal, Jasmin Quandt, Daniel Maurus, et al. Covid-19 vaccine bnt162b1 elicits human antibody and th 1 t cell responses. *Nature*, 586(7830):594–599, 2020.

Karsten Sauer and Tim Harris. An effective COVID-19 vaccine needs to engage T cells. *Frontiers in Immunology*, 11, 2020.

Kathryn E Stephenson, Mathieu Le Gars, Jerald Sadoff, Anne Marit De Groot, Dirk Heerwagh, Carla Truyers, Caroline Atyeo, Carolin Loos, Abishek Chandrashekhar, Katherine McMahan, et al. Immunogenicity of the ad26. cov2. s vaccine for covid-19. *Jama*, 325(15):1535–1544, 2021.

Phillip A Swanson, Marcelino Padilla, Wesley Hoyland, Kelly McGlinchey, Paul A Fields, Sagida Bibi, Saul N Faust, Adrian B McDermott, Teresa Lambe, Andrew J Pollard, et al. T-cell mediated immunity after azd1222 vaccination: A polyfunctional spike-specific th1 response with a diverse tcr repertoire. *medRxiv*, 2021.

Gianluca Tomasello, Ilaria Armenia, and Gianluca Molla. The Protein Imager: a full-featured online molecular viewer interface with server-side HQ-rendering capabilities. *Bioinformatics*, 36(9):2909–2911, 2020.

WHO. WHO coronavirus (COVID-19) dashboard. <https://covid19.who.int/>, 2021. Accessed: 2021-08-31.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4.
URL <https://ggplot2.tidyverse.org>.

Hadley Wickham. *tidyverse: Tidy Messy Data*, 2020. URL <https://CRAN.R-project.org/package=tidyr>. R package version 1.1.2.

Hadley Wickham, Romain François, Lionel Henry, and Kirill Müller. *dplyr: A Grammar of Data Manipulation*, 2020. URL <https://CRAN.R-project.org/package=dplyr>. R package version 1.0.2.