Methods

TCR data

This study is based on public data from three previous works (Alter et al., 2021; Nolan et al., 2020; Mayer-Blackwell et al., 2020).

The dataset used for TCR repertoire analysis (Alter et al., 2021) 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 TCRβ 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™ database (Nolan et al., 2020). The full dataset contains TCR repertoires from 193 convalescent patients whose blood sample was collected at the Hospital Univesitario 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$^+$ TCRβ sequences were obtained from Mayer-Blackwell et al. (2020). This sequences are proven to bind SARS-CoV-2 epitopes by Multiplex Identification of Receptor Antigen (MIRA) assays (Nolan et al., 2020) and are also enriched in bulk TCRβ repertoires of convalescent individuals compared to healthy controls. For the present study, only TCRβ 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 (Alter et al., 2021; Gittelman et al., 2021), those enriched and high-reliable datasets are not currently public. ImmuneCODE™ database 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. The enrichment analysis was performed with tcrdist3 Python toolkit (Docker image v0.1.9) (Mayer-Blackwell et al., 2020; Dash et al., 2017), following the same meta-clonotype discovery pipeline employed for SARS-CoV-2 CD8$^+$ TCR discovery as in Mayer-Blackwell et al. (2020).

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

The 43 TCR repertoires were annotated for antigen-specificity with the SARS-CoV-2-specific TCRs (CD4$^+$ and CD8$^+$) by matching 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 (Mayer-Blackwell et al., 2020; Dash et al., 2017), 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 ☐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 ≤ 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 (Csardi and Nepusz, 2006).

Data analysis and visualization

All plots and analyses were carried in R 3.6.1 (R Core Team, 2019). For data analysis, the packages dplyr v1.0.2 (Wickham et al., 2020), tidyr v1.1.2 (Wickham, 2020), rstatix v0.7.0 (Kassambara, 2021) and parallel v3.6.1 (R Core Team, 2019) were used. Networks were plotted with the ggraph v2.0.2 package (Pedersen, 2020). All other plots were generated with ggplot2 v3.3.2 (Wickham, 2016) and ggpubr v0.4.0 (Kassambara, 2020).

Bibliography

Galit Alter, Jingyou Yu, Jinyan Liu, Abishek Chandrashekar, 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 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.

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.

Alboukadel Kassambara. *ggpubr: 'ggplot2' Based Publication Ready Plots*, 2020. URL `https://CRAN.R-project.org/package=ggpubr`. 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.

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.

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 (TCRβ) 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/`.

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. *tidyr: 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.