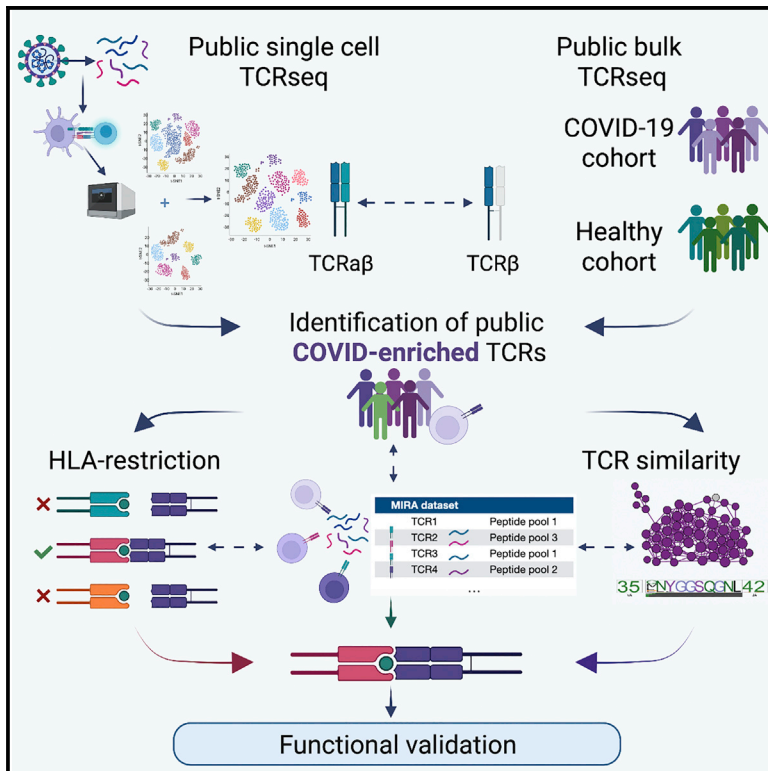


# Resolving SARS-CoV-2 CD4<sup>+</sup> T cell specificity via reverse epitope discovery

## Graphical abstract



## Authors

Mikhail V. Pogorelyy, Elisa Rosati, Anastasia A. Minervina, ..., Andre Franke, Petra Bacher, Paul G. Thomas

## Correspondence

p.bacher@ikmb.uni-kiel.de (P.B.), paul.thomas@stjude.org (P.G.T.)

## In brief

Pogorelyy et al. perform a meta-analysis of large single-cell and bulk T cell receptor (TCR) datasets from SARS-CoV-2-infected individuals to identify public CD4<sup>+</sup> responses. This study reports more than 1,200  $\alpha\beta$ TCRs forming six prominent similarity clusters and validates HLA restriction and epitope-specificity predictions for five clusters using transgenic T cell lines.

## Highlights

- Identification and experimental validation of CD4<sup>+</sup> T cell responses to SARS-CoV-2
- Prediction of HLA restriction and epitopes for CD4<sup>+</sup> SARS-CoV-2-specific responses
- General reverse epitope discovery framework, applicable to other diseases



## Report

# Resolving SARS-CoV-2 CD4<sup>+</sup> T cell specificity via reverse epitope discovery

Mikhail V. Pogorelyy,<sup>1,4</sup> Elisa Rosati,<sup>2,3,4</sup> Anastasia A. Minervina,<sup>1,4</sup> Robert C. Mettelman,<sup>1</sup> Alexander Scheffold,<sup>3</sup> Andre Franke,<sup>2</sup> Petra Bacher,<sup>2,3,\*</sup> and Paul G. Thomas<sup>1,5,\*</sup>

<sup>1</sup>Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN 38103, USA

<sup>2</sup>Institute of Clinical Molecular Biology, Christian-Albrecht University of Kiel, Kiel, Germany

<sup>3</sup>Institute of Immunology, Christian-Albrecht University of Kiel, Kiel, Germany

<sup>4</sup>These authors contributed equally

<sup>5</sup>Lead contact

\*Correspondence: [p.bacher@ikmb.uni-kiel.de](mailto:p.bacher@ikmb.uni-kiel.de) (P.B.), [paul.thomas@stjude.org](mailto:paul.thomas@stjude.org) (P.G.T.)

<https://doi.org/10.1016/j.xcrm.2022.100697>

## SUMMARY

The current strategy to detect immunodominant T cell responses focuses on the antigen, employing large peptide pools to screen for functional cell activation. However, these approaches are labor and sample intensive and scale poorly with increasing size of the pathogen peptidome. T cell receptors (TCRs) recognizing the same epitope frequently have highly similar sequences, and thus, the presence of large sequence similarity clusters in the TCR repertoire likely identify the most public and immunodominant responses. Here, we perform a meta-analysis of large, publicly available single-cell and bulk TCR datasets from severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-infected individuals to identify public CD4<sup>+</sup> responses. We report more than 1,200  $\alpha\beta$ TCRs forming six prominent similarity clusters and validate histocompatibility leukocyte antigen (HLA) restriction and epitope specificity predictions for five clusters using transgenic T cell lines. Collectively, these data provide information on immunodominant CD4<sup>+</sup> T cell responses to SARS-CoV-2 and demonstrate the utility of the reverse epitope discovery approach.

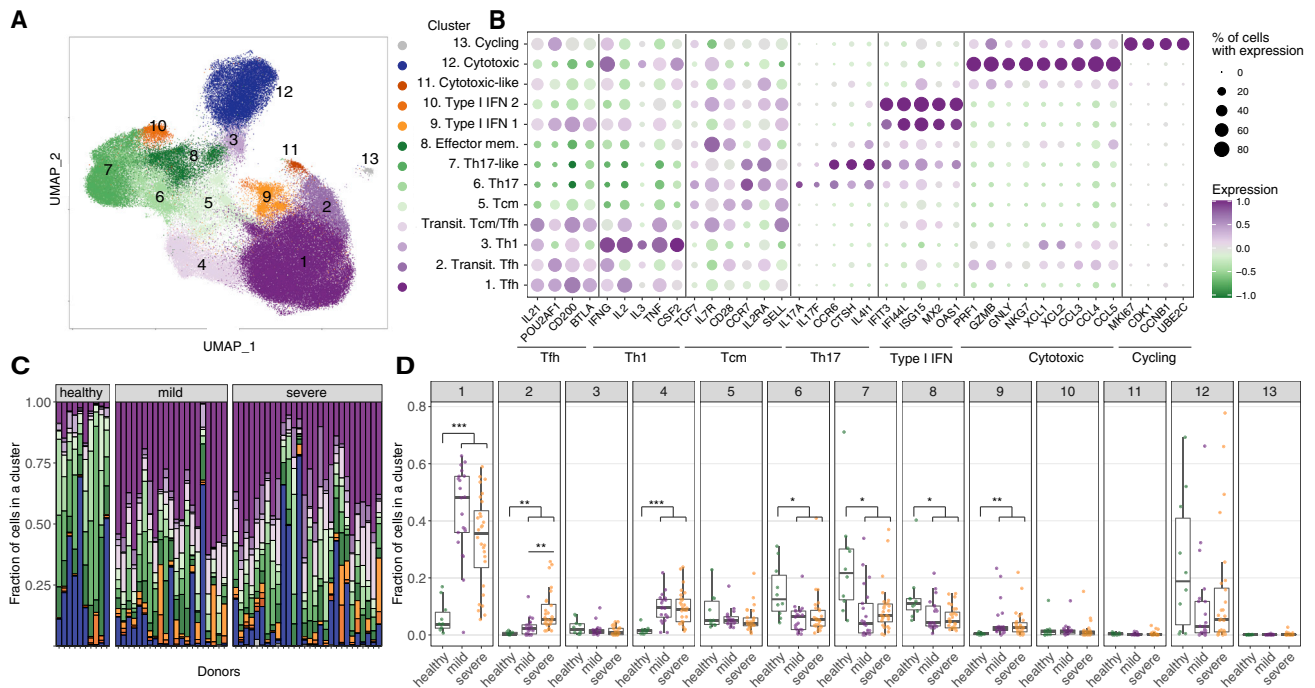
## INTRODUCTION

The global scientific effort to overcome the COVID-19 pandemic has led to the generation of an extraordinarily large amount of publicly available data describing the human immune response to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Many of these studies point to the importance of robust T cell responses in resolving COVID-19, as well as providing long-term protection against newly emerging SARS-CoV-2 variants.<sup>6–13</sup> One increasingly popular method to study T cell response complexity is to characterize the T cell receptor (TCR) repertoires of activated T cells following infection or vaccination.<sup>14</sup> The TCR is a heterodimer of alpha and beta chains, both of which are formed in a semi-random DNA recombination process resulting in a unique repertoire in each individual that can be resolved through sequencing.<sup>15</sup> However, despite their complex and unique nature, TCR repertoires often have similar and definable features across individuals encountering the same antigens. In particular, responses to immunodominant epitopes trigger large clonal expansions, and TCRs recognizing such epitopes frequently have highly similar sequences.<sup>16–18</sup> Thus, analysis of TCR repertoires could shed light on the differences and commonalities of the T cell immune response across individuals, help discern the identities of the most dominant immunogenic antigens, and provide targets for development of diagnostic and therapeutic strategies, such as adoptive T cell

transfer. Indeed, identification of T cell clonotypes reactive to SARS-CoV-2 antigens has led to the development of a novel diagnostic test for SARS-CoV-2 authorized for emergency use by the US Food and Drug Administration (FDA).<sup>19</sup>

In order to identify TCR repertoire signatures related to COVID-19, multiple groups have utilized either bulk TCR repertoire sequencing methods,<sup>3,20–24</sup> which quantitatively measure the frequencies of large numbers of unpaired TCR alpha or beta chain clonotypes, or single-cell TCR-sequencing techniques,<sup>1,2,25–31</sup> which produce fewer but paired alpha/beta TCR sequences. One of the major challenges of TCR-sequencing approaches is that only a small fraction of total peripheral T cells recognize viral epitopes, even at the peak of anti-SARS-CoV-2 immune responses. Hence, many studies rely on methods designed to enrich antigen-specific T cell responses prior to TCR sequencing, such as the antigen-reactive T cell enrichment strategy (ARTE), activation-induced marker (AIM), and multiplex identification of T cell receptor antigen specificity (MIRA) assays,<sup>32–35</sup> which combine major histocompatibility complex (MHC)-multimer staining or peptide stimulation with subsequent selection of activated T cells. These methods increase the number of SARS-CoV-2-specific TCRs detected in each sample and help to identify immunodominant epitopes (reviewed in Grifoni et al.<sup>36</sup>). Stimulating T cells with peptide libraries is the most frequently used approach in SARS-CoV-2 epitope discovery.<sup>28,37–42</sup> However, peptide stimulation across





**Figure 1. Merged analysis of single-cell CD4<sup>+</sup> SARS-CoV-2-reactive T cell public datasets**

(A) Uniform Manifold Approximation and Projection (UMAP) of single cells from merged datasets containing SARS-CoV-2-antigen-enriched CD4 T cells. Colors indicate clusters of cells with distinct gene expression profiles.

(B) Differentially expressed genes in each gene expression (GEX) cluster.

(C) Distribution of cells between GEX clusters is plotted for each donor; clusters of healthy donors do not contain Tfh cells (populations 1 and 2).

(D) Boxplots depicting the fraction of cells among functional clusters for each participant (Mann-Whitney U test; Bonferroni multiple comparison correction, \*p < 0.05, \*\*p < 0.005, \*\*\*p < 0.0005).

the whole viral peptidome requires very large sample sizes, which becomes increasingly challenging and cost prohibitive for bigger pathogens, such as SARS-CoV-2. Moreover, instead of a uniform response against all available peptides, the immune system tends to focus on several so-called “immunodominant” epitopes that we are yet unable to predict based on epitope features alone. Here, we propose a reverse epitope discovery technique, which moves the burden of T cell epitope detection away from large peptide screens and rather utilizes rich TCR repertoire datasets as the means to identify unbiased immunodominant responses. We performed a comprehensive TCR meta-analysis of publicly available single-cell and bulk CD4<sup>+</sup> TCR repertoire datasets and identified more than 1,200 highly public SARS-CoV-2-reactive TCRs with complete TCR alpha and beta chain information and inferred their corresponding HLA restriction. Moreover, by clustering TCRs based on sequence similarity, we were able to (1) identify several prominent alpha/beta TCR motifs, (2) predict their antigen specificity, and (3) validate the prediction experimentally using transgenic T cell lines, demonstrating the utility of the reverse epitope discovery approach.

## RESULTS

In this report, we jointly analyzed multiple datasets across three study groupings. The first grouping focused on two published single-cell datasets of SARS-CoV-2-reactive CD4<sup>+</sup> T cells

identified based on CD154<sup>+</sup> up-regulation after peptide pool stimulation,<sup>33</sup> with associated gene expression and TCR information from a total of 59 individuals (49 COVID-19 positive; 10 healthy unexposed unvaccinated controls).<sup>1,2</sup> The second grouping utilized the largest published bulk TCRbeta datasets currently available, which together comprise 786 healthy pre-pandemic samples<sup>4</sup> and 1,414 COVID-19 patients<sup>3</sup> and include a TCR sub-dataset (MIRA class II dataset) with known specificity for certain SARS-CoV-2 peptide pools.<sup>22</sup> Lastly, our third group, which was used to validate our findings, included a published bulk TCR dataset obtained from SARS-CoV-2-unexposed participants sampled before and after immunization with the AstraZeneca *ChAdOx1* SARS-CoV-2 vaccine.<sup>5</sup>

In order to find public CD4<sup>+</sup> T cell responses to SARS-CoV-2 infection, we first merged two publicly available single-cell datasets of CD4<sup>+</sup> SARS-CoV-2-reactive T cells.<sup>1,2</sup> Both datasets were obtained using the same antigen-reactive T cell enrichment strategy (ARTE assay). Of note, in the study from Bacher et al.,<sup>1</sup> the peptide pools span the spike, membrane, and nucleocapsid proteins, while peptide pools used in Meckiff et al.<sup>2</sup> only include peptides from the spike (without N-terminal domain) and membrane proteins. The combined dataset contained 125,258 cells that passed quality control steps, which resulted in 13 functional clusters after unsupervised analysis (Figure 1A). Cluster phenotypes were defined using cell-population markers used in the original publications.<sup>1,2</sup> In particular, we found clusters

corresponding to T follicular helper (Tfh) cells (clusters 1 and 2), type 1 T helper (Th1) cells (cluster 3), transitional Tfh and T central memory (Tcm) cells (cluster 4), fully differentiated Tcm cells (cluster 5), Th17 phenotypes (clusters 6 and 7), effector memory T (Tem) cells (cluster 8), type I interferon (IFN)-signature T cells (clusters 9 and 10), cytotoxic T cells (clusters 11 and 12), and cycling T cells (cluster 13) (Figures 1A and 1B; Table S1). Using this extended dataset, we also confirmed findings reported in the study from Meckiff et al.<sup>2</sup> of significant Tfh enrichment (clusters 1 and 2) in COVID-19 patients compared with unexposed healthy controls (Figures 1C and 1D). Interestingly, the abundance of the two main Tfh subsets was significantly different in hospitalized versus non-hospitalized COVID-19 patients (Figure 1D). In particular, cluster 2 was significantly enriched in severe disease ( $p = 0.004$ ; after Benjamini-Hochberg multiple testing correction) and expressed higher cytotoxic markers, such as *CCL3*, *CCL4*, *CCL5*, *XCL1*, *XCL2*, *GZMB*, and *GNLY*. In contrast, cluster 1, expressing high levels of *IL-2* and *CD69* and genes of the *TNF* family, was more prevalent in patients with mild disease, as previously shown in Meckiff et al.,<sup>2</sup> although not reaching statistical significance in the merged dataset ( $p = 0.4$  after Benjamini-Hochberg multiple testing correction) (Figure 1D; Table S1).

In order to select TCRs corresponding to the most public CD4<sup>+</sup> T cell responses, we next searched TCRbeta sequences from the combined TCR alpha/beta single-cell dataset in the TCRbeta repertoires from a large cohort of COVID-19 patients,<sup>3</sup> as well as from pre-pandemic COVID-19-naive controls.<sup>4</sup> We identified 1,248 unique alpha/beta TCRs shared among individuals that are strongly enriched in COVID-19 patients as compared with controls (COVID-enriched TCRs) ( $p < 0.05$ ; after Benjamini-Hochberg correction; Fisher's exact test), which are reported in Table S2. We also identified 594 alpha/beta TCRs, which were significantly decreased in COVID-19 patients as compared with the healthy pre-pandemic controls (COVID-depleted TCRs), reported in Table S3 (Figure 2A). Notably, when mapping COVID-enriched TCRs to the single-cell RNA sequencing (RNA-seq) data, these TCRs significantly accumulated in Tfh-containing clusters (clusters 1, 2, and 4), while COVID-depleted TCRs, on the other hand, accumulated in Tem subpopulations (clusters 6–8) (Figure 2B). The overall effect size (log<sub>2</sub>-fold enrichment) for COVID-depleted TCRs was much smaller than for COVID-enriched TCRs (Figure 2A). Moreover, COVID-depleted TCRs were present in a large fraction of donors from both the control and the COVID-19 cohorts (Figure 2C); in particular, 465 out of 594 were simultaneously found in >100 controls and in >100 COVID-19 patients. We hypothesize these COVID-depleted clonotypes are a consequence of COVID-19-associated lymphopenia. In fact, the number of unique T cell clones in a subset of the analyzed COVID-19 patients was lower in comparison to healthy controls. This could lead to small yet significant underrepresentation of highly public clonotypes in the COVID-19 cohort. We therefore focused on the COVID-enriched TCR clonotypes for further analysis, as the occurrence pattern and phenotype of this group is consistent with expansion of T cell clones specific for SARS-CoV-2 antigens.

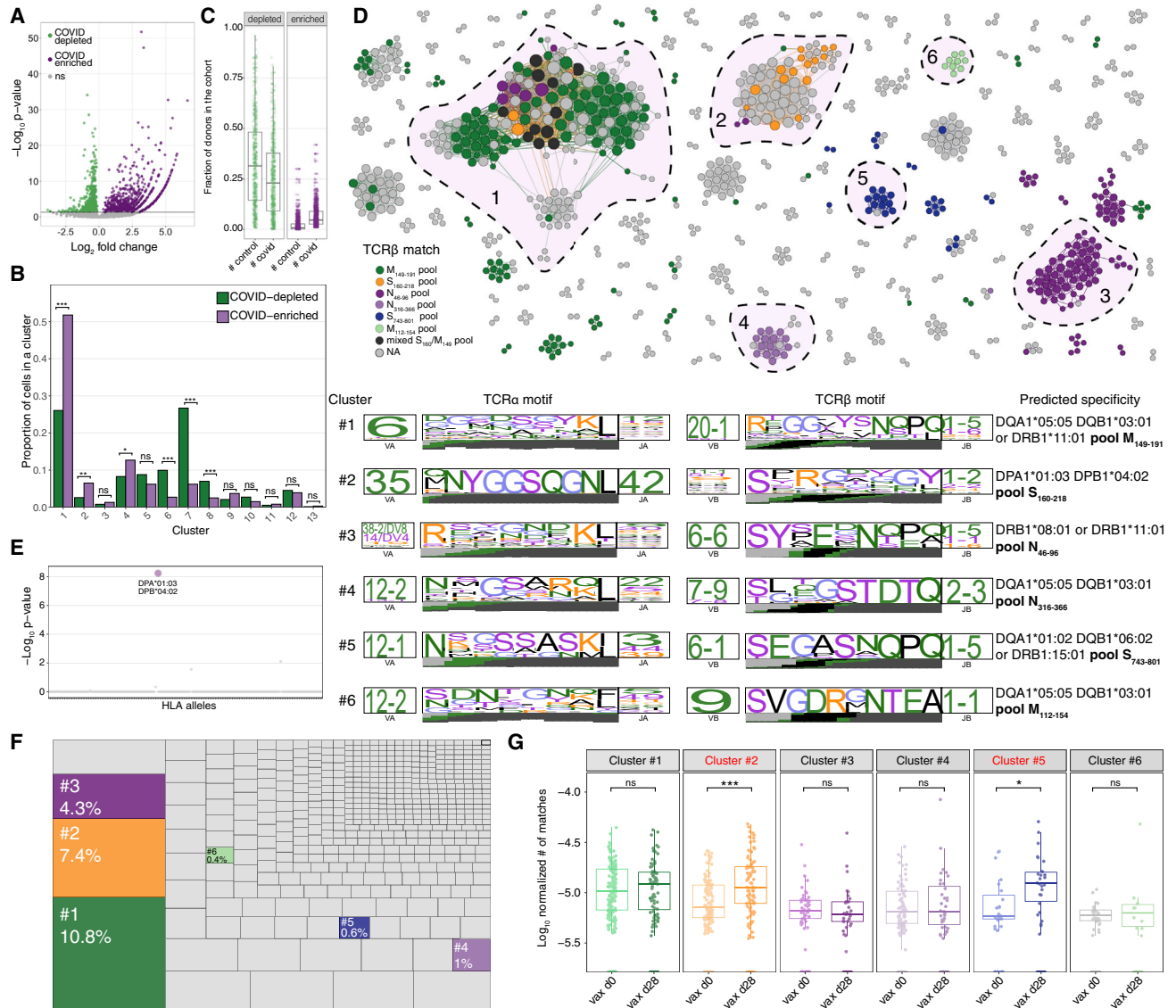
In order to assess the sequence similarity among TCRs enriched in COVID-19, we used TCRdist,<sup>16</sup> which calculates a

TCR sequence distance (similarity) measure optimized for clustering highly similar TCRs often with the same specificity. This analysis showed the presence of several prominent TCR clusters (Figure 2D). Interestingly, cluster 3 is largely defined by a conserved beta chain motif, allowing for diverse alpha chains, while in cluster 2, there is an almost invariant alpha chain paired with a set of very diverse TCRbeta chains. In a few other large clusters, both TCR chains show strongly conserved amino acid motifs (Figure 2D). These differences could be potentially explained by the variable number of contacts of TCR alpha/beta chains with the antigenic peptide and MHC. Thus, alpha-driven, beta-driven, and alpha/beta-driven motifs are interesting targets for solving TCR-peptide MHC (pMHC) ternary structures.

As a TCR only binds its cognate epitope presented in the context of a specific HLA molecule, an individual's HLA background can strongly influence the composition of the TCR repertoire. Thus, people with overlapping HLA alleles will have more similar TCR repertoires than people with different HLA backgrounds. This feature of the TCR repertoire was successfully exploited in an elegant study by De-Witt et al.,<sup>43</sup> where the authors were able to identify a set of TCR sequences associated with certain HLA alleles. Using this set of HLA-associated TCRs, we inferred the HLA types of the COVID-19 patients within the bulk TCR dataset,<sup>3</sup> in which conventional HLA typing information was not available. Therefore, to derive associations between TCR clonotypes and HLA, we first determined whether participants with specific COVID-19-enriched TCR clonotypes carried a common HLA allele when compared with TCR-negative patients. Using this approach, we were then able to predict a potential HLA association of our COVID-enriched TCR set using Fisher's exact test to evaluate enrichment within different HLA contexts. See Figure 2E for the representative Manhattan plot and Table S2 for predicted HLA restrictions.

To predict the potential antigen reactivity of the TCR clusters, we next cross-referenced the COVID-enriched TCRbeta sequences with the MIRA MHC class II dataset.<sup>22</sup> This dataset contains TCRbeta sequences of CD4<sup>+</sup> T cells with known specificity to 1 of 56 peptide pools containing one to six overlapping 19-mer peptides spanning the membrane (M), nucleocapsid (N), and spike (S) SARS-CoV-2 proteins. Based on TCR similarity, 428 of 1,242 (34%) COVID-enriched TCRs were successfully mapped to 22 of 56 peptide pools from the MIRA dataset, with most matches to the M<sub>149–191</sub>, N<sub>46–96</sub>, and S<sub>743–801</sub> protein pools. Most TCRs belonging to the same TCRdist cluster were assigned to the same peptide pool. Interestingly, the biggest TCRdist cluster (cluster 1) included mostly TCRs mapping to the membrane protein pool M<sub>149–191</sub>; however, a few TCRs in this cluster were assigned to multiple pools and proteins of the MIRA database, which we hypothesize to be a minor confounding effect of the antigen-reactivity deconvolution approach of the MIRA assay. To further narrow down the specificity prediction to the epitope level, we used NetMHCIIpan4.0 to identify potential binding peptides within SARS-CoV-2 S, M, and N regions covered by the predicted MIRA peptide pools and restricted by HLA alleles determined for each cluster. Potential binders were identified within each region. Large TCR similarity clusters—those containing high





**Figure 2. Reverse epitope discovery of SARS-CoV-2-reactive public CD4<sup>+</sup> T cell clonotypes**

(A) Volcano plot shows enrichment of TCRbeta chains from merged single-cell TCR sequencing (scTCR-seq) datasets in a large (n = 1,414) collection of bulk TCRbeta repertoires from COVID-19 patients<sup>3</sup> (purple) in comparison to the healthy donor cohort from Emerson et al.<sup>4</sup> (n = 786) (x axis) versus p value (y axis). ns, not significant.

(B) Barplot showing the distribution of COVID-enriched (purple) and COVID-depleted (green) TCR clonotypes in GEX clusters. Fisher's exact test was used for the comparison, with Bonferroni multiple comparison correction. \*p < 0.05, \*\*p < 0.005, \*\*\*p < 0.0005, ns, not significant.

(C) The boxplots show the fraction of donors from healthy and COVID-19 cohorts sharing significantly COVID-depleted (green) and COVID-enriched (purple) clonotypes.

(D) A similarity network of COVID-associated public TCR clonotypes. Each vertex represents a TCR alpha/beta clonotype, and edges connect vertices with <120 TCRdist units. Colors show predicted specificity to SARS-CoV-2 peptide pools from the MIRA class II dataset. Bottom: TCRdist logos for the most prominent clonotype clusters with predicted peptide specificity and HLA restriction are shown.

(E) Manhattan plot for association of representative clonotypes from cluster 2 with various HLA types.

(F) A tree map showing the fraction of T cells of the merged single-cell dataset carrying clonotypes from the prominent TCR similarity clusters from (C).

(G) Occurrence of TCRbeta from six large clusters from (C) prior to and following SARS-CoV-2 vaccination with the ChAdOx1 (AstraZeneca) vaccine. Significantly more TCRs from spike specific (clusters 2 and 5) are found after vaccination (one-sided Wilcoxon rank-sum test with Benjamini-Hochberg multiple testing correction).

numbers of unique clonotypes found in many different individuals—are likely to recognize immunodominant epitopes. Indeed, the three TCR similarity clusters with the greatest

number of TCR clones also correspond to the largest magnitude of response when accounting for the clone size (Figure 2F).

To further validate the set of COVID-enriched CD4<sup>+</sup> clonotypes using an independent dataset, we used a large collection of TCRbeta repertoires from healthy unexposed individuals before and after immunization with *ChAdOx1*,<sup>5</sup> a replication-deficient simian adenovirus-vectored vaccine encoding the SARS-CoV-2 S protein produced by AstraZeneca. For each individual, we calculated the fraction of unique TCRbeta clonotypes identified in our largest antigen-specific TCR clusters out of total clonotypes sampled at either the pre-vaccination (day 0) or post-vaccination (day 28) time point (Figure 2D). As expected, only TCRs reactive to S protein antigens (clusters 2 and 5) were significantly enriched after vaccination, while the frequency of TCR clusters reactive to the M (1 and 6) or N proteins (3 and 4) remained unchanged (Figure 2G). The same significant enrichment was observed when analyzing all TCR clusters predicted to be reactive for the S protein (Figure S1A). Interestingly, at least one of the prominent orphan TCR clusters (defined here as TCR clusters without a match in the MIRA dataset and thus unknown antigen reactivity) also showed significant enrichment after vaccination, suggesting TCR reactivity to the S protein (Figure S1B). This result serves as an independent validation that our reverse epitope discovery approach correctly predicts antigen reactivity, at least at the protein level, and at the same time demonstrates how the TCR clusters we resolved may be used to identify SARS-CoV-2 epitope-specific TCRs. Excitingly, as we correctly identified epitopes within the context of both vaccination and natural infection, this approach may have the potential to distinguish between these two conditions and, further, to define the relative contribution of external and internal proteins to the overall individual immune response.

Next, we sought to experimentally validate our predicted pMHC epitopes using assays to probe pMHC and TCR engagement and evaluate functional activation. To do this, we transduced TCR-null Jurkat cells with constructs encoding TCRs with representative TCR sequences from five of the six largest TCR clusters (Figure 2D). These individual Jurkat-TCR cell lines, now expressing the encoded TCRs at the cell surface, were co-cultured with peripheral blood mononuclear cells (PBMCs) from healthy donors carrying the predicted HLA restriction necessary to bind and present the peptide epitopes. The co-cultured cells were then pulsed with overlapping 17-mer peptides spanning the predicted MIRA peptide pools and evaluated for TCR activation using the endogenous NFAT-GFP reporter in the transgenic Jurkat cells (Figure 3A). All generated T cell lines reacted to one or two specific peptides (Figure 3B), confirming our specificity prediction and allowing us to precisely determine the epitope location within a MIRA pool. Importantly, for four of five TCRs, the experimentally determined reactive peptide region overlaps with the peptide region identified to be a binder by NetMHC within the context of a specific MIRA pool and HLA allele. Moreover, the predicted antigenic peptide and HLA restriction for TCRdist cluster 2 and cluster 5 also exactly match the experimental results from two independent studies.<sup>28,44</sup> However, for cluster 6, the predicted immunogenic region did not match the experimental results, likely due to a misassigned HLA restriction. To test this, we compared NetMHC binding estimates for the reactive peptides across all HLA class II alleles present on the donor PBMCs (Figure 3C). While, for four out of five

TCRs, the highest binding affinity towards the experimentally defined cognate peptide was in the context of the predicted HLAs, for cluster 6, the HLA allele predicted to have the highest binding to the identified reactive peptides was DRB3\*02:02, rather than the predicted allele, DQB1\*03:01/DQA1\*05:05. Importantly, DRB3 allele information is absent in many standard HLA-typing datasets and in the reference dataset<sup>43</sup> utilized by us and thus could not be predicted by our approach.

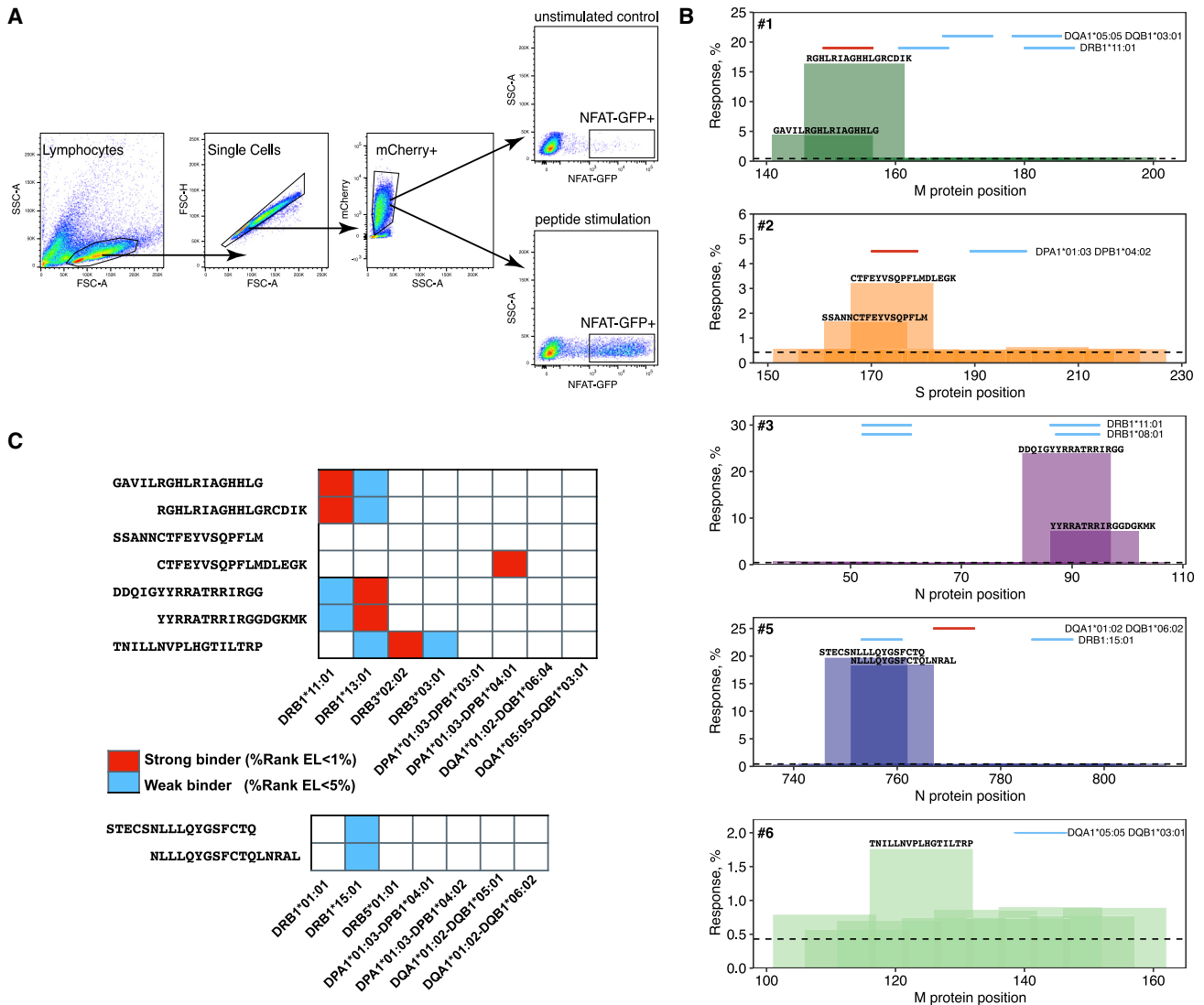
Together, our data demonstrate the utility of using T cell repertoire and individual specificity to predict antigen binding down to single epitope resolution. Although the quality of these predictions depends on the limitations of the source datasets, it is clear that this method is a viable alternative to peptide-focused epitope discovery approaches and will only improve as more datasets are published. Future expansion of public resources, including TCR repertoire sequencing data, will further improve TCR specificity prediction accuracy and may also allow the identification of certain TCR repertoire features correlated with protection from SARS-CoV-2 infection and/or severe disease, such as a high frequency of SARS-CoV-2-specific clones or sharing of certain public clonotypes recognizing immunodominant epitopes.

## DISCUSSION

Here, we describe an approach to identify public immunodominant CD4<sup>+</sup> T cell responses based on TCR amino acid sequence similarity. Combined with other datasets and methods, this approach may help to resolve new epitopes across different immune response contexts. As a proof of concept, we identified 1,248 paired TCR clonotypes potentially specific to highly immunogenic epitopes from SARS-CoV-2. Many of these were so-called orphan TCRs without defined epitopes. However, using an approach we term reverse epitope discovery, we successfully identified and inferred antigen specificity for 428 TCRs with the aid of the MIRA dataset.<sup>22</sup> We also inferred possible HLA restrictions for most of these TCRs (88%). TCR-HLA pairings were further validated based on NetMHC binding predictions and functional experiments in this study and by others.<sup>28,44</sup> The resulting set of highly characterized public TCRs reactive to SARS-CoV-2 covers more than 76% of individuals from Snyder et al.,<sup>3</sup> indicating that at least 20 unique COVID-enriched TCR sequences per individual were found using this approach. Interestingly, two out of six major responses were restricted by HLA alleles with low variability among the human population (DPB1-04:01:02 and DRB3-02:02), thus providing a means to further investigate immunodominant responses in genetically diverse individuals. Furthermore, the high publicity of the characterized clonotypes makes them promising candidates for further studies on CD4<sup>+</sup> T cell immune responses mounted against SARS-CoV-2 as well as for immunotherapeutic applications aimed at utilizing highly specific, immunodominant T cell responses in the context of precision and personalized medicine.<sup>45–47</sup>

## Limitations of the study

The described method is strongly focused on public T cell responses, which provide the necessary power for a robust statistical analysis and, further, hold the highest potential for



**Figure 3. Results of TCR specificity validation experiment**

(A) Gating strategy. Jurkat activation is tracked by GFP expression under NFAT control.

(B) Peptides triggering the response for each analyzed cell line are shown on the corresponding regions of SARS-CoV-2 M, N, and S proteins. The height of the bars indicates the percentage of antigen-specific response of the NFAT-GFP TCR transgenic Jurkat cell lines in co-culture with PBMCs from healthy donors pulsed with overlapping 17-mer peptides covering the predicted antigenic region. Dashed lines show the background activation level of the corresponding transgenic Jurkat cell line in an unstimulated sample. The previously computationally predicted epitope and HLA are indicated above each plot in blue or red lines for weak and strong predicted HLA binders, respectively.

(C) HLA restriction prediction by NetMHC: the identified immunogenic peptides are computationally tested for HLA binding against the HLA alleles present on the donor PBMCs used in the experiment; colors show peptides overlapping with weak (blue) and strong (red) HLA-binding cores.

additional population-wide applications. However, in many cases, immunodominant responses within a particular individual may be driven by private clonotypes or clonotypes without an identifiable motif cluster, which would be missed by the current approach or would not appear among the most interesting hits. Applying different sequence similarity measures and integrating more datasets will likely expand the number of identified T cell responses with this approach. The major limitation for the TCR-repertoire-based HLA typing algorithm utilized here is the co-inheritance of certain HLA alleles. Since the algorithm is

based on the co-occurrence of a TCR sequence with a particular HLA allele, linked HLA alleles from a haplotype can be difficult to distinguish. However, even narrowing down the possible HLA restriction to several HLA alleles can significantly facilitate TCR specificity determination. HLA typing based on TCR repertoire data is also challenging for rare, non-classical, or low-diversity alleles (e.g., DRB3/4/5), as no known TCRs have been associated with these loci to date. More datasets, including deeply sequenced TCR repertoires from HLA-typed individuals can address this limitation in the future.

While the detection of TCR clusters relies on bulk and/or single-cell TCR sequencing, peptide target prediction requires more sophisticated datasets containing TCR antigen-specificity information, such as VDjdb<sup>48</sup> and the MIRA dataset.<sup>22</sup> The incompletely overlapping peptide pools utilized for our reference single-cell datasets can limit the reverse epitope discovery approach. For example, cluster 2 from Figure 2 is formed mostly by clonotypes from Bacher et al.<sup>1</sup> and absent in Meckiff et al.<sup>2</sup> Indeed, the peptide pools used in the latter did not cover the N-terminal domain of the S protein and thus did not include the target peptide of the TCRs in cluster 2. Although this is a limitation of these datasets, it also indirectly confirms the predicted antigen specificity of cluster 2. The expansion of these resources is of the utmost importance for target identification of orphan TCRs. The strong general interest in the COVID-19 pandemic led to an unprecedented production of high amounts of publicly available TCR repertoire data, which still far exceed data available for other antigens or diseases.

Despite these limitations, the reverse epitope discovery approach described here has already proven valuable for identifying both cross-reactive<sup>49</sup> and immunodominant responses to SARS-CoV-2,<sup>44</sup> and it holds potential for application in other disease contexts.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
  - Lead contact
  - Materials availability
  - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Cell lines
  - Utilized public data
- **METHOD DETAILS**
  - Transgenic T cell lines generation
  - Peptide stimulation assay
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - Single-cell datasets integration and filtering
  - COVID-19 TCR association using bulk TCR public datasets
  - Identification of motifs in TCR amino acid sequences using TCRdist
  - HLA specificity imputation from TCR data
  - Prediction of COVID-enriched TCR specificity
  - Statistical analysis

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2022.100697>.

## ACKNOWLEDGMENTS

This research was supported by the German Research Foundation (DFG) under Germany's Excellence Strategy—EXC 2167-390884018 Precision Medi-

cine in Chronic Inflammation to P.B., A.F., and A.S.; DFG 433038070 to P.B., A.F., and A.S.; and DFG 4096610003 to E.R.—and by a COVID-19 research grant from the Land Schleswig-Holstein, DIO002/CoVispecT to P.B. and A.S. This research was partially supported by R01AI136514 (P.G.T.) and Ruth L. Kirschstein National Research Service Award (NRSA) Individual Postdoctoral Fellowship (NRSA-NIAID) award number F32AI157296 (R.C.M.).

## AUTHOR CONTRIBUTIONS

Conceptualization, E.R., M.V.P., and A.A.M.; analysis, E.R., M.V.P., A.A.M., and R.C.M.; visualization, A.A.M. and E.R.; resources, P.G.T., P.B., A.S., and A.F.; supervision and coordination, P.B. and P.G.T.; writing—original draft, E.R., M.V.P., and A.A.M. All authors provided discussion, participated in revising the manuscript, and agreed to the final version.

## DECLARATION OF INTERESTS

P.B. and A.S. are consultants of Miltenyi Biotec, who own IP rights concerning parts of the ARTE technology. P.G.T. has consulted and/or received honoraria and travel support from Illumina, Johnson and Johnson, and 10X Genomics. P.G.T. serves on the Scientific Advisory Board of Immunoscope and Cytoagents. The authors have applied for patents covering some aspects of these studies.

Received: November 18, 2021

Revised: May 8, 2022

Accepted: June 24, 2022

Published: July 1, 2022

## REFERENCES

1. Bacher, P., Rosati, E., Esser, D., Martini, G.R., Saggau, C., Schiminsky, E., Dargvainiene, J., Schröder, I., Wieters, I., Khodamoradi, Y., et al. (2020). Low-avidity CD4+ T cell responses to SARS-CoV-2 in unexposed individuals and humans with severe COVID-19. *Immunity* 53, 1258–1271.e5. <https://doi.org/10.1016/j.immuni.2020.11.016>.
2. Meckiff, B.J., Ramírez-Suástegui, C., Fajardo, V., Chee, S.J., Kusnadi, A., Simon, H., Eschweiler, S., Grifoni, A., Pelosi, E., Weiskopf, D., et al. (2020). Imbalance of regulatory and cytotoxic SARS-CoV-2-reactive CD4+ T cells in COVID-19. *Cell* 183, 1340–1353.e16. <https://doi.org/10.1016/j.cell.2020.10.001>.
3. Snyder, T.M., Gittelman, R.M., Klinger, M., May, D.H., Osborne, E.J., Taniguchi, R., Zahid, H.J., Kaplan, I.M., Dines, J.N., Noakes, M.T., et al. (2020). Magnitude and dynamics of the T-cell response to SARS-CoV-2 infection at both individual and population levels. Preprint at medRxiv. <https://doi.org/10.1101/2020.07.31.20165647>.
4. Emerson, R.O., DeWitt, W.S., Vignali, M., Gravley, J., Hu, J.K., Osborne, E.J., Desmarais, C., Klinger, M., Carlson, C.S., Hansen, J.A., et al. (2017). Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat. Genet.* 49, 659–665. <https://doi.org/10.1038/ng.3822>.
5. Swanson, P.A., Padilla, M., Hoyland, W., McGlinchey, K., Fields, P.A., Bibi, S., Faust, S.N., McDermott, A.B., Lambe, T., Pollard, A.J., et al.; AstraZeneca/Oxford/VRC Study Group (2021). AZD1222/ChAdOx1 nCoV-19 vaccination induces a polyfunctional spike protein-specific Th1 response with a diverse TCR repertoire. *Sci. Transl. Med.* 13, eabj7211. <https://doi.org/10.1126/scitranslmed.abj7211>.
6. Dan, J.M., Mateus, J., Kato, Y., Hastie, K.M., Yu, E.D., Faliti, C.E., Grifoni, A., Ramirez, S.I., Haupt, S., Frazier, A., et al. (2021). Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* 371, eabf4063. <https://doi.org/10.1126/science.abf4063>.
7. Geers, D., Shamier, M.C., Bogers, S., den Hartog, G., Gommers, L., Nieuwkoop, N.N., Schmitz, K.S., Rijsbergen, L.C., van Osch, J.A.T., Dijkhuizen, E., et al. (2021). SARS-CoV-2 variants of concern partially escape



- humoral but not T-cell responses in COVID-19 convalescent donors and vaccinees. *Sci. Immunol.* 6, eabj1750. <https://doi.org/10.1126/sciimmunol.abj1750>.
8. Kundu, R., Narean, J.S., Wang, L., Fenn, J., Pillay, T., Fernandez, N.D., Conibear, E., Koycheva, A., Davies, M., Tolosa-Wright, M., et al. (2022). Cross-reactive memory T cells associate with protection against SARS-CoV-2 infection in COVID-19 contacts. *Nat. Commun.* 13, 80. <https://doi.org/10.1038/s41467-021-27674-x>.
  9. Peng, Y., Felce, S.L., Dong, D., Penkava, F., Mentzer, A.J., Yao, X., Liu, G., Yin, Z., Chen, J.L., Lu, Y., et al. (2022). An immunodominant NP105–113-B\*07:02 cytotoxic T cell response controls viral replication and is associated with less severe COVID-19 disease. *Nat. Immunol.* 23, 50–61. <https://doi.org/10.1038/s41590-021-01084-z>.
  10. Rydzynski Moderbacher, C., Ramirez, S.I., Dan, J.M., Grifoni, A., Hastie, K.M., Weiskopf, D., Belanger, S., Abbott, R.K., Kim, C., Choi, J., et al. (2020). Antigen-specific adaptive immunity to SARS-CoV-2 in acute COVID-19 and associations with age and disease severity. *Cell* 183, 996–1012.e19. <https://doi.org/10.1016/j.cell.2020.09.038>.
  11. Sekine, T., Perez-Potti, A., Rivera-Ballesteros, O., Stralin, K., Gorin, J.B., Olsson, A., Llewellyn-Lacey, S., Kamal, H., Bogdanovic, G., Muschiol, S., et al. (2020). Robust T cell immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell* 183, 158–168.e14. <https://doi.org/10.1016/j.cell.2020.08.017>.
  12. Swadling, L., Diniz, M.O., Schmidt, N.M., Amin, O.E., Chandran, A., Shaw, E., Pade, C., Gibbons, J.M., Le Bert, N., Tan, A.T., Jeffery-Smith, A., Tan, C.C.S., Tham, C.Y.L., Kucykowicz, S., Aidoo-Micah, G., Rosenheim, J., Davies, J., Johnson, M., Jensen, M.P., Joy, G., McCoy, L.E., Valdes, A.M., Chain, B.M., Goldblatt, D., Altmann, D.M., Boyton, R.J., Manisty, C., Treibel, T.A., Moon, J.C., COVIDsortium Investigators; van Dorp, L., Balloux, F., McKnight, Á., Noursadeghi, M., Bertoletti, A., and Maini, M.K. (2022). Pre-existing polymerase-specific T cells expand in abortive seronegative SARS-CoV-2. *Nature* 601, 110–117. <https://doi.org/10.1038/s41586-021-04186-8>.
  13. Tarke, A., Sidney, J., Methot, N., Yu, E.D., Zhang, Y., Dan, J.M., Goodwin, B., Rubiro, P., Sutherland, A., Wang, E., et al. (2021). Impact of SARS-CoV-2 variants on the total CD4+ and CD8+ T cell reactivity in infected or vaccinated individuals. *Cell Rep. Med.* 2, 100355. <https://doi.org/10.1016/j.xcrm.2021.100355>.
  14. Mukhopadhyay, M. (2021). Diving into the TCR repertoire. *Nat. Methods* 18, 30. <https://doi.org/10.1038/s41592-020-01031-0>.
  15. Dupic, T., Bensouda Koraichi, M., Minervina, A.A., Pogorelyy, M.V., Mora, T., and Walczak, A.M. (2021). Immune fingerprinting through repertoire similarity. *PLoS Genet.* 17, e1009301. <https://doi.org/10.1371/journal.pgen.1009301>.
  16. Dash, P., Fiore-Gartland, A.J., Hertz, T., Wang, G.C., Sharma, S., Souquette, A., Crawford, J.C., Clemens, E.B., Nguyen, T.H.O., Kedzierska, K., et al. (2017). Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature* 547, 89–93. <https://doi.org/10.1038/nature22383>.
  17. Glanville, J., Huang, H., Nau, A., Hatton, O., Wagar, L.E., Rubelt, F., Ji, X., Han, A., Krams, S.M., Pettus, C., et al. (2017). Identifying specificity groups in the T cell receptor repertoire. *Nature* 547, 94–98. <https://doi.org/10.1038/nature22976>.
  18. Pogorelyy, M.V., Minervina, A.A., Shugay, M., Chudakov, D.M., Lebedev, Y.B., Mora, T., and Walczak, A.M. (2019). Detecting T cell receptors involved in immune responses from single repertoire snapshots. *PLoS Biol.* 17, e3000314. <https://doi.org/10.1371/journal.pbio.3000314>.
  19. Dalai, S.C., Dines, J.N., Snyder, T.M., Gittelman, R.M., Eerkes, T., Vaney, P., Howard, S., Akers, K., Skewis, L., Monteforte, A., et al. (2021). Clinical validation of a novel T-cell receptor sequencing assay for identification of recent or prior SARS-CoV-2 infection. Preprint at medRxiv. <https://doi.org/10.1101/2021.01.06.21249345>.
  20. Minervina, A.A., Komech, E.A., Titov, A., Bensouda Koraichi, M., Rosati, E., Mamedov, I.Z., Franke, A., Efimov, G.A., Chudakov, D.M., Mora, T., et al. (2021). Longitudinal high-throughput TCR repertoire profiling reveals the dynamics of T-cell memory formation after mild COVID-19 infection. *Elife* 10, e63502. <https://doi.org/10.7554/elife.63502>.
  21. Niu, X., Li, S., Li, P., Pan, W., Wang, Q., Feng, Y., Mo, X., Yan, Q., Ye, X., Luo, J., et al. (2020). Longitudinal analysis of T and B cell receptor repertoire transcripts reveal dynamic immune response in COVID-19 patients. *Front. Immunol.* 11, 582010. <https://doi.org/10.3389/fimmu.2020.582010>.
  22. Nolan, S., Vignali, M., Klinger, M., Dines, J.N., Kaplan, I.M., Svejnova, E., Craft, T., Boland, K., Pesesky, M., Gittelman, R.M., et al. (2020). A large-scale database of T-cell receptor beta (TCR $\beta$ ) sequences and binding associations from natural and synthetic exposure to SARS-CoV-2. *Res. Square*. <https://doi.org/10.21203/rs.3.rs-51964/v1>.
  23. Schultheiß, C., Paschold, L., Simnica, D., Mohme, M., Willscher, E., von Wenserski, L., Scholz, R., Wieters, I., Dahlke, C., Tolosa, E., et al. (2020). Next-generation sequencing of T and B cell receptor repertoires from COVID-19 patients showed signatures associated with severity of disease. *Immunity* 53, 442–455.e4. <https://doi.org/10.1016/j.immuni.2020.06.024>.
  24. Shomradova, A.S., Vagida, M.S., Sheetikov, S.A., Zornikova, K.V., Kiryukhin, D., Titov, A., Peshkova, I.O., Khmelevskaya, A., Dianov, D.V., Malasheva, M., et al. (2020). SARS-CoV-2 epitopes are recognized by a public and diverse repertoire of human T cell receptors. *Immunity* 53, 1245–1257.e5. <https://doi.org/10.1016/j.immuni.2020.11.004>.
  25. Bernardes, J.P., Mishra, N., Tran, F., Bahmer, T., Best, L., Blase, J.I., Bordoni, D., Franzenburg, J., Geisen, U., Josephs-Spauding, J., et al.; HCA Lung Biological Network (2020). Longitudinal multi-omics analyses identify responses of megakaryocytes, erythroid cells, and plasmablasts as hallmarks of severe COVID-19. *Immunity* 53, 1296–1314.e9. <https://doi.org/10.1016/j.immuni.2020.11.017>.
  26. Kusnadi, A., Ramírez-Suástegui, C., Fajardo, V., Chee, S.J., Meckiff, B.J., Simon, H., Pelosi, E., Seumois, G., Ay, F., Vijayanand, P., and Ottensmeier, C.H. (2021). Severely ill COVID-19 patients display impaired exhaustion features in SARS-CoV-2-reactive CD8<sup>+</sup> T cells. *Sci. Immunol.* 6, eabe4782. <https://doi.org/10.1126/sciimmunol.abe4782>.
  27. Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat. Med.* 26, 842–844. <https://doi.org/10.1038/s41591-020-0901-9>.
  28. Lu, X., Hosono, Y., Nagae, M., Ishizuka, S., Ishikawa, E., Motooka, D., Ozaki, Y., Sax, N., Maeda, Y., Kato, Y., et al. (2021). Identification of conserved SARS-CoV-2 spike epitopes that expand public cTfh clonotypes in mild COVID-19 patients. *J. Exp. Med.* 218, e20211327. <https://doi.org/10.1084/jem.20211327>.
  29. Wen, W., Su, W., Tang, H., Le, W., Zhang, X., Zheng, Y., Liu, X., Xie, L., Li, J., Ye, J., et al. (2020). Author Correction: immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing. *Cell Discov.* 6, 41. <https://doi.org/10.1038/s41421-020-00187-5>.
  30. Xu, G., Qi, F., Li, H., Yang, Q., Wang, H., Wang, X., Liu, X., Zhao, J., Liao, X., Liu, Y., et al. (2020). The differential immune responses to COVID-19 in peripheral and lung revealed by single-cell RNA sequencing. *Cell Discov.* 6, 73. <https://doi.org/10.1038/s41421-020-00225-2>.
  31. Zhang, J.-Y., Wang, X.M., Xing, X., Xu, Z., Zhang, C., Song, J.W., Fan, X., Xia, P., Fu, J.L., Wang, S.Y., et al. (2020). Single-cell landscape of immunological responses in patients with COVID-19. *Nat. Immunol.* 21, 1107–1118. <https://doi.org/10.1038/s41590-020-0762-x>.
  32. Altman, J.D., Moss, P.A.H., Goulder, P.J.R., Barouch, D.H., McHeyzer-Williams, M.G., Bell, J.I., McMichael, A.J., and Davis, M.M. (1996). Phenotypic analysis of antigen-specific T lymphocytes. *Science* 274, 94–96. <https://doi.org/10.1126/science.274.5284.94>.
  33. Bacher, P., Schink, C., Teutschbein, J., Kniemeyer, O., Assenmacher, M., Brakhage, A.A., and Scheffold, A. (2013). Antigen-reactive T cell enrichment for direct, high-resolution analysis of the human naive and memory Th cell repertoire. *J. Immunol.* 190, 3967–3976. <https://doi.org/10.4049/jimmunol.1202221>.

34. Klinger, M., Pepin, F., Wilkins, J., Asbury, T., Wittkop, T., Zheng, J., Moorhead, M., and Faham, M. (2015). Multiplex identification of antigen-specific T cell receptors using a combination of immune assays and immune receptor sequencing. *PLoS One* 10, e0141561. <https://doi.org/10.1371/journal.pone.0141561>.
35. Reiss, S., Baxter, A.E., Cirelli, K.M., Dan, J.M., Morou, A., Daigneault, A., Brassard, N., Silvestri, G., Routy, J.P., Havenar-Daughton, C., et al. (2017). Comparative analysis of activation induced marker (AIM) assays for sensitive identification of antigen-specific CD4 T cells. *PLoS One* 12, e0186998. <https://doi.org/10.1371/journal.pone.0186998>.
36. Grifoni, A., Sidney, J., Vita, R., Peters, B., Crotty, S., Weiskopf, D., and Sette, A. (2021). SARS-CoV-2 Human T cell Epitopes: adaptive immune response against COVID-19. *Cell Host Microbe* 29, 1076–1092. <https://doi.org/10.1016/j.chom.2021.05.010>.
37. Braun, J., Loyal, L., Frentsch, M., Wendisch, D., Georg, P., Kurth, F., Hippenstiel, S., Dingeldey, M., Kruse, B., Fauchere, F., et al. (2020). SARS-CoV-2-reactive T cells in healthy donors and patients with COVID-19. *Nature* 587, 270–274. <https://doi.org/10.1038/s41586-020-2598-9>.
38. Grifoni, A., Weiskopf, D., Ramirez, S.I., Mateus, J., Dan, J.M., Moderbacher, C.R., Rawlings, S.A., Sutherland, A., Premkumar, L., Jadi, R.S., et al. (2020). Targets of T Cell responses to SARS-CoV-2 coronavirus in humans with COVID-19 disease and unexposed individuals. *Cell* 181, 1489–1501.e15. <https://doi.org/10.1016/j.cell.2020.05.015>.
39. Mateus, J., Grifoni, A., Tarke, A., Sidney, J., Ramirez, S.I., Dan, J.M., Burger, Z.C., Rawlings, S.A., Smith, D.M., Phillips, E., et al. (2020). Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* 370, 89–94. <https://doi.org/10.1126/science.abd3871>.
40. Nelde, A., Bilich, T., Heitmann, J.S., Maringer, Y., Salihi, H.R., Roerden, M., Lübke, M., Bauer, J., Rieth, J., Wacker, M., et al. (2021). SARS-CoV-2 derived peptides define heterologous and COVID-19-induced T cell recognition. *Nat. Immunol.* 22, 74–85. <https://doi.org/10.1038/s41590-020-00808-x>.
41. Peng, Y., Mentzer, A.J., Liu, G., Yao, X., Yin, Z., Dong, D., Dejnirattisai, W., Rostron, T., Supasa, P., Liu, C., et al. (2020). Broad and strong memory CD4+ and CD8+ T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nat. Immunol.* 21, 1336–1345. <https://doi.org/10.1038/s41590-020-0782-6>.
42. Tarke, A., Sidney, J., Kidd, C.K., Dan, J.M., Ramirez, S.I., Yu, E.D., Mateus, J., da Silva Antunes, R., Moore, E., Rubiro, P., et al. (2021). Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. *Cell Rep. Med.* 2, 100204. <https://doi.org/10.1016/j.xcrm.2021.100204>.
43. DeWitt, W.S., Smith, A., Schoch, G., Hansen, J.A., Matsen, F.A., and Bradley, P. (2018). Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *Elife* 7, e38358. <https://doi.org/10.7554/elife.38358>.
44. Mudd, P.A., Minervina, A.A., Pogorelyy, M.V., Turner, J.S., Kim, W., Kalaidina, E., Petersen, J., Schmitz, A.J., Lei, T., Haile, A., et al. (2022). SARS-CoV-2 mRNA vaccination elicits a robust and persistent T follicular helper cell response in humans. *Cell* 185, 603–613.e15. <https://doi.org/10.1016/j.cell.2021.12.026>.
45. Müller, T.R., Jarosch, S., Hammel, M., Leube, J., Grassmann, S., Bernard, B., Effenberger, M., Andrä, I., Chaudhry, M.Z., Käuferle, T., et al. (2021). Targeted T cell receptor gene editing provides predictable T cell product function for immunotherapy. *Cell Rep. Med.* 2, 100374. <https://doi.org/10.1016/j.xcrm.2021.100374>.
46. O'Reilly, R.J., Prockop, S., Hasan, A.N., Koehne, G., and Doubrovina, E. (2016). Virus-specific T-cell banks for 'off the shelf' adoptive therapy of refractory infections. *Bone Marrow Transplant.* 51, 1163–1172. <https://doi.org/10.1038/bmt.2016.17>.
47. Qian, C., Wang, Y., Reppel, L., D'aveni, M., Campidelli, A., Decot, V., and Bensoussan, D. (2018). Viral-specific T-cell transfer from HSCT donor for the treatment of viral infections or diseases after HSCT. *Bone Marrow Transplant.* 53, 114–122. <https://doi.org/10.1038/bmt.2017.232>.
48. Bagaev, D.V., Vroomans, R.M.A., Samir, J., Stervbo, U., Rius, C., Dolton, G., Greenshields-Watson, A., Attaf, M., Egorov, E.S., Zvyagin, I.V., et al. (2020). VDjdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. *Nucleic Acids Res.* 48, D1057–D1062. <https://doi.org/10.1093/nar/gkz874>.
49. Minervina, A.A., Pogorelyy, M.V., Kirk, A.M., Crawford, J.C., Allen, E.K., Chou, C.H., Mettelman, R.C., Allison, K.J., Lin, C.Y., Brice, D.C., et al. (2022). SARS-CoV-2 antigen exposure history shapes phenotypes and specificity of memory CD8+ T cells. *Nat. Immunol.* 23, 781–790. <https://doi.org/10.1038/s41590-022-01184-4>.
50. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* 36 (5), 411–420. <https://doi.org/10.1038/nbt.4096>.
51. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* 16, 1289–1296. <https://doi.org/10.1038/s41592-019-0619-0>.
52. Shugay, M., Britanova, O.V., Merzlyak, E.M., Turchaninova, M.A., Mamedov, I.Z., Tuganbaev, T.R., et al. (2014). Towards error-free profiling of immune repertoires. *Nature Methods* 11 (6), 653–655. <https://doi.org/10.1038/nmeth.2960>.
53. Bolotin, D. A., Poslavsky, S., Mitrophanov, I., Shugay, M., Mamedov, I. Z., Putintseva, E. V., & Chudakov, D. M. MiXCR: Software for comprehensive adaptive immunity profiling. *Nature Methods* 2015, 12(5), 380–381. <https://doi.org/10.1038/nmeth.3364>
54. Schattgen, S.A., Guion, K., Crawford, J.C., Souquette, A., Barrio, A.M., Stubbington, M.J.T., Thomas, P.G., and Bradley, P. (2021). Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). *Nat. Biotechnol.* 40, 54–63. <https://doi.org/10.1038/s41587-021-00989-2>.
55. Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Int. J. Complex Syst.* 1695, 1–9.
56. Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS One* 9, e98679. <https://doi.org/10.1371/journal.pone.0098679>.
57. Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Pric, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol.* 16, 278. <https://doi.org/10.1186/s13059-015-0844-5>.
58. Reynisson, B., Alvarez, B., Paul, S., Peters, B., and Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.* 48, W449–W454. <https://doi.org/10.1093/nar/gkaa379>.

## STAR★METHODS

### KEY RESOURCES TABLE

| REAGENT or RESOURCE                                                                                                                                     | SOURCE                         | IDENTIFIER                                                                                                                                            |
|---------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Antibodies</b>                                                                                                                                       |                                |                                                                                                                                                       |
| Co-stimulatory anti-human CD28 antibody (clone CD28.2)                                                                                                  | BD Biosciences                 | cat#: 555725 RRID: AB_396068                                                                                                                          |
| Co-stimulatory anti-human CD49d antibody (clone 9F10)                                                                                                   | BD Biosciences                 | cat#: 555501 RRID: AB_2130052                                                                                                                         |
| <b>Chemicals, peptides, and recombinant proteins</b>                                                                                                    |                                |                                                                                                                                                       |
| Lenti-X Concentrator                                                                                                                                    | Clontech                       | cat#: 631232                                                                                                                                          |
| 1x Cell Stimulation cocktail                                                                                                                            | eBioscience                    | cat#: 00-4970-93                                                                                                                                      |
| SARS-CoV-2 peptides (>95% purity)                                                                                                                       | This paper                     | Genbank acc: MT019529.1                                                                                                                               |
| <b>Deposited data</b>                                                                                                                                   |                                |                                                                                                                                                       |
| Single-cell RNA-seq of SARS-CoV-2-reactive CD4 <sup>+</sup> T cells                                                                                     | Bacher et al. <sup>1</sup>     | SRA: SRP293741                                                                                                                                        |
| Single-cell RNA-seq of SARS-CoV-2-reactive CD4 <sup>+</sup> T cells                                                                                     | Meckiff et al. <sup>2</sup>    | SRA: SRP267404                                                                                                                                        |
| Bulk TCR repertoire from SARS-CoV-2-infected individuals                                                                                                | Snyder et al. <sup>3</sup>     | <a href="https://clients.adaptivebiotech.com/pub/covid-2020">https://clients.adaptivebiotech.com/pub/covid-2020</a>                                   |
| Bulk TCR repertoire from healthy individuals                                                                                                            | Emerson et al. <sup>4</sup>    | <a href="https://clients.adaptivebiotech.com/pub/emerson-2017-natgen">https://clients.adaptivebiotech.com/pub/emerson-2017-natgen</a>                 |
| Bulk TCR repertoire from SARS-CoV-2 unexposed participants sampled before and after immunization with the AstraZeneca <i>ChAdOx1</i> SARS-CoV-2 vaccine | Swanson et al. <sup>5</sup>    | <a href="https://clients.adaptivebiotech.com/pub/emerson-2017-natgen">https://clients.adaptivebiotech.com/pub/emerson-2017-natgen</a>                 |
| MIRA class II bulk TCR dataset with known specificity for certain SARS-CoV-2 peptide pools (release 002.1)                                              | Nolan et al. <sup>22</sup>     | <a href="https://clients.adaptivebiotech.com/pub/covid-2020">https://clients.adaptivebiotech.com/pub/covid-2020</a>                                   |
| Original code for data processing                                                                                                                       | This paper                     | <a href="https://github.com/pogorely/reverse_epitope_discovery">https://github.com/pogorely/reverse_epitope_discovery</a>                             |
| <b>Experimental models: Cell lines</b>                                                                                                                  |                                |                                                                                                                                                       |
| 293T                                                                                                                                                    | ATCC                           | cat#:CRL-3216                                                                                                                                         |
| Jurkat 76.7 (variant of TCR-null Jurkat 76.7 cells that expresses human CD8 and an NFAT-GFP reporter)                                                   | gift from Wouter Schepers      |                                                                                                                                                       |
| <b>Recombinant DNA</b>                                                                                                                                  |                                |                                                                                                                                                       |
| pLVX-EF1 $\alpha$ -IRES-Puro                                                                                                                            | Clontech                       | cat#: 631253                                                                                                                                          |
| TCR_cluster1-mCherry                                                                                                                                    | This paper                     |                                                                                                                                                       |
| TCR_cluster2-mCherry                                                                                                                                    | This paper                     |                                                                                                                                                       |
| TCR_cluster3-mCherry                                                                                                                                    | This paper                     |                                                                                                                                                       |
| TCR_cluster5-mCherry                                                                                                                                    | This paper                     |                                                                                                                                                       |
| TCR_cluster6-mCherry                                                                                                                                    | This paper                     |                                                                                                                                                       |
| psPAX2 packaging plasmid                                                                                                                                | gift from Didier Trono         | Addgene plasmid #12260<br>RRID: Addgene_12260                                                                                                         |
| pMD2.G envelope plasmid                                                                                                                                 | gift from Didier Trono         | Addgene plasmid #12259<br>RRID: Addgene_12259                                                                                                         |
| <b>Software and algorithms</b>                                                                                                                          |                                |                                                                                                                                                       |
| FlowJo v10.7.1                                                                                                                                          | BD Biosciences                 | <a href="https://www.flowjo.com/solutions/flowjo/downloads">https://www.flowjo.com/solutions/flowjo/downloads</a>                                     |
| Cell Ranger v3.1.0                                                                                                                                      | 10x Genomics                   | <a href="https://www.10xgenomics.com">https://www.10xgenomics.com</a>                                                                                 |
| Seurat v.3.2.0                                                                                                                                          | Butler et al. <sup>50</sup>    | <a href="https://www.satijalab.org/seurat">https://www.satijalab.org/seurat</a>                                                                       |
| Harmony v1.0                                                                                                                                            | Korsunsky et al. <sup>51</sup> | <a href="https://portals.broadinstitute.org/harmony/articles/quickstart.html">https://portals.broadinstitute.org/harmony/articles/quickstart.html</a> |
| R v. 4.0.2                                                                                                                                              |                                | <a href="https://www.r-project.org">https://www.r-project.org</a>                                                                                     |
| Biorender                                                                                                                                               |                                | <a href="https://biorender.com">https://biorender.com</a>                                                                                             |
| MiGEC v. 1.2.7                                                                                                                                          | Shugay et al. <sup>52</sup>    | <a href="https://github.com/mikessh/migec">https://github.com/mikessh/migec</a>                                                                       |
| MiXCR v. 3.0.3                                                                                                                                          | Bolotin et al. <sup>53</sup>   | <a href="https://github.com/milaboratory/mixcr">https://github.com/milaboratory/mixcr</a>                                                             |

(Continued on next page)

**Continued**

| REAGENT or RESOURCE             | SOURCE                          | IDENTIFIER                                                                                                                          |
|---------------------------------|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------|
| CoNGA python package            | Schattgen et al. <sup>54</sup>  | <a href="https://github.com/phbradley/conga">https://github.com/phbradley/conga</a>                                                 |
| data.table R package v. 1.14.0  |                                 | <a href="https://github.com/Rdatatable/data.table/wiki">https://github.com/Rdatatable/data.table/wiki</a>                           |
| stringdist R package v. 0.9.6.3 |                                 | <a href="https://github.com/markvanderloo/stringdist">https://github.com/markvanderloo/stringdist</a>                               |
| igraph R package v. 1.2.6       | Csardi and Nepusz <sup>55</sup> | <a href="https://igraph.org/r/">https://igraph.org/r/</a>                                                                           |
| gephi v. 0.9.2                  | Jacomy et al. <sup>56</sup>     | <a href="https://gephi.org">https://gephi.org</a>                                                                                   |
| ggplot2 R package v. 3.3.3      |                                 | <a href="https://cran.r-project.org/web/packages/ggplot2/index.html">https://cran.r-project.org/web/packages/ggplot2/index.html</a> |

**RESOURCE AVAILABILITY**

**Lead contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Paul G. Thomas ([paul.thomas@stjude.org](mailto:paul.thomas@stjude.org)).

**Materials availability**

All transgenic T cell lines generated in this study are available from the [lead contact](#) with a completed Materials Transfer Agreement. No other unique reagents were generated in this study.

**Data and code availability**

Single-cell data can be found under SRA accession numbers SRA: SRP293741 for ref. <sup>1</sup> and SRA: SRP267404 for ref. <sup>2</sup>. Bulk TCR repertoire data from SARS-CoV-2-infected subjects, and the MIRA Class II dataset (release 002.1)<sup>3</sup> are publicly available from the ImmuneAccess database (<https://clients.adaptivebiotech.com/pub/covid-2020>). Data obtained from healthy participants<sup>4</sup> can be located in (<https://clients.adaptivebiotech.com/pub/emerson-2017-natgen>) and in (<https://doi.org/10.21417/PAS2021STM>) for ChAdOx1 immunized cohort dataset.<sup>5</sup>

Original code used to analyze merged single-cell datasets are available on the GitHub repository [https://github.com/pogorely/reverse\\_epitope\\_discovery](https://github.com/pogorely/reverse_epitope_discovery).

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

**EXPERIMENTAL MODEL AND SUBJECT DETAILS**

**Cell lines**

TCR-null Jurkat 76.7 cell lines with endogenous NFAT-GFP reporter were generously provided by Wouter Scheper and were maintained in RPMI (Gibco) containing 10% FBS and 1% penicillin/streptomycin.

**Utilized public data**

Single-cell data of SARS-CoV-2 reactive CD154<sup>+</sup> T cells were obtained from refs. <sup>1</sup> and <sup>2</sup>. Bulk TCR data of healthy individuals and COVID-19 patients were obtained from ref. <sup>4</sup> and ref. <sup>3</sup>, respectively.

**METHOD DETAILS**

**Transgenic T cell lines generation**

For the six most prominent TCR similarity clusters from [Figure 2](#), we selected a representative TCR sequence (the sequence with the most number of neighbors) for cloning into a TCR-null Jurkat 76.7 cell line (generously provided by Wouter Scheper), which express an NFAT-GFP reporter. TCRalpha and TCRbeta sequences were altered to use murine constant regions (murine TRAC\*01 and murine TRBC2\*01) to facilitate surface expression of the TCR. Six gene fragments were synthesized by Genscript to encode TCRalpha, TCRbeta chain, and mCherry fluorescent protein, linked together by 2A sites and put into the pLVX-EF1 $\alpha$ -IRES-Puro lentiviral backbone (Clontech). To generate transducing particles, HEK 293T packaging cells (ATCC CRL-3216) were transfected with an individual TCR-encoding lentiviral vector, psPAX2 packaging plasmid (Addgene plasmid #12260), and pMD2.G envelope plasmid (Addgene plasmid #12259). We collected transducing particle-containing media 24- and 48-h post-transfection and concentrated the lentivirus with Lenti-X Concentrator (Clontech) according to the manufacturer's protocol. Jurkat 76.7 cells were transduced, then antibiotic selected for 1 week using 1  $\mu$ g/mL puromycin in RPMI (Gibco) containing 10% FBS and 1% penicillin/streptomycin. Five out of six Jurkat cell lines were successfully transduced as confirmed by expression of mCherry.



### Peptide stimulation assay

Jurkat cell line expressing select TCRs ( $10^5$ ) were co-cultured with PBMCs from pre-pandemic healthy donors ( $2 \times 10^5$ ) in cRPMI, pulsed with 1  $\mu$ M of peptide, and co-stimulated with 1  $\mu$ g/mL each of anti-human CD28 and CD49d (BD Biosciences). Negative unstimulated (no peptide; CD28, CD49d) and positive (CD28, CD49d, 1X Cell Stimulation Cocktail, PMA/ionomycin; eBioscience) controls were included in each assay. Cells were incubated for 18 h (37°C, 5% CO<sub>2</sub>). After incubation cells were washed twice with FACS buffer (PBS, 2% FBS, 1 mM EDTA), and analyzed by flow cytometry on a custom-configured BD Fortessa using FACSDiva software (Becton Dickinson). Flow cytometry data were analyzed using FlowJo software (BD Biosciences). Responsiveness to peptide stimulation was determined by measuring the frequency of cells positive for endogenous NFAT-GFP expression (Table S4).

## QUANTIFICATION AND STATISTICAL ANALYSIS

### Single-cell datasets integration and filtering

Preprocessing of scRNAseq data was performed with 10x Genomics' Cell Ranger software v3.1.0 using the human genome reference GRCh38 v3.0.0 for the mapping. The resulting raw feature-barcode matrix files were analyzed with the Seurat v.3.2.0 R package (Butler et al., 2018). All genes with detected expression in less than 0.1% of the cells were excluded. TCR genes were not considered for further analyses to avoid functional clustering of cells based on TCR information. For cell quality control only cells harboring between 400 and 3000 RNA features and less than 5% mitochondrial RNA were selected for further processing.

TCR information was integrated into the Seurat object metadata after filtering cells containing more than 2 TCRalpha or 2 TCRbeta chains. After merging of Seurat metadata and TCR information, cells without TCR information were excluded from further analysis. Afterwards, data were log-normalized and scaled based on all genes. After performing a PCA dimensionality reduction (40 dimensions) with the RunPCA function, expression values were corrected for batch effects caused by different sources of the data, sample preparation batches, and sequencing run batches using the R package Harmony v1.0.<sup>51</sup> In the final steps, the Uniform Manifold Approximation and Projection (UMAP) dimensional reduction was performed with the RunUMAP function using 40 dimensions, a shared nearest neighbor graph was created with the FindNeighbors method, and the clusters identification was performed with a resolution of 0.4 using the FindClusters function. 13 clusters were identified. Cluster marker genes were determined using FindMarkers with the MAST method<sup>57</sup> and are available in Table S1.

### COVID-19 TCR association using bulk TCR public datasets

To identify public TCRbeta clonotypes we used two large datasets, one of TCRbeta repertoires from COVID-19 patients ( $n = 1414$ )<sup>3</sup> and one of healthy subjects sampled pre-pandemic ( $n = 786$ ).<sup>4</sup> For each TCRbeta from the combined single-cell TCRseq dataset we calculated the number of unique donors from the both bulk TCRbeta repertoire cohorts sharing the clonotype. TCRbeta clonotypes were considered shared if both CDR3 amino acid sequences and V segment families matched. Next, we use a two-sided Fisher's exact test with Benjamini-Hochberg multiple-comparisons correction to identify overrepresented (i.e. found in more donors) TCRbeta clonotypes in either COVID-19 or healthy donors (adjusted  $p$ -value < 0.05 is used as significance threshold).

### Identification of motifs in TCR amino acid sequences using TCRdist

We used the TCRdist implementation in the CoNGA python package to calculate pairwise TCRdist between unique alpha/betaTCR sequences and to plot sequence logos for TCR motifs.<sup>54</sup> We define TCR motifs as a cluster on the TCR similarity network, where each node is a unique alpha/betaTCR clonotype, and edges connect nodes if the distance between them is less than 120 TCRdist units. To filter TCR chimeras and other artifacts occurring during 10x Genomics sequencing leading to rare spurious connections between TCR motif clusters, we deleted the top 1% of nodes and vertices by network betweenness centrality values. The *igraph* R package was used to manipulate similarity networks,<sup>55</sup> *gephi* was used for network layout and visualization.<sup>56</sup>

### HLA specificity imputation from TCR data

For each donor from refs.<sup>3,4</sup> we use HLA-types inferred as previously described in ref.<sup>20</sup> In brief, for each TCRbeta significantly enriched in the COVID-19 cohort we performed a one-sided Fisher's exact test with Benjamini-Hochberg multiple-comparisons correction to check if a given TCRbeta co-occurs with a certain HLA-allele. To determine an HLA restriction within a TCR similarity cluster we considered the HLA with the most significant association for each individual TCR in the cluster. If the HLA-allele was the best prediction for 25% or more TCRs of the cluster, it was considered associated with that cluster.

### Prediction of COVID-enriched TCR specificity

We mapped the TCRbeta chain sequences from the merged aggregated single-cell dataset to peptide pool-specific TCRbeta clonotypes of the MIRA class II dataset (release 002.1) allowing for one amino acid mismatch between CDR3 amino acid sequences. Next we selected six large clusters on the TCRdist similarity network with distinct MIRA peptide pool assignments. The cluster was assigned a certain MIRA pool if at least 20% of TCR sequences within a cluster mapped to a MIRA pool; it was considered "orphan" otherwise. We calculated the consensus MIRA pool and HLA-restriction within each TCR cluster. We then used NetMHCIIpan-4.0<sup>58</sup> to predict the epitope location within the MIRA pool, for the specific HLA-restriction. We ran NetMHCIIpan-4.0 on the complete

SARS-CoV-2 Wuhan-1 S, M, N protein sequences (Genbank acc: MT019529.1) for peptide lengths 9–15, for predicted HLAs (see [Figure 2D](#)), all other parameters were set to default. We then checked for presence of strongly ( $\text{Rank\_EL} < 1\%$ ) and weakly ( $\text{Rank\_EL} < 5\%$ ) HLA-binding peptide cores within the predicted MIRA region. To confirm the HLA-restriction of responding peptides in the validation experiments, we used the same method with slight modifications: HLA-binding peptide cores were first predicted for each of the antigen-presenting cell donors' HLA alleles, after which we checked for overlap of these binding cores within peptides eliciting a response in the cloned TCR-expressing Jurkat cell lines (see [Figure 3C](#)).

### Statistical analysis

Statistical analyses were performed in R version 4.0.2. Wilcoxon rank-sum test (Mann-Whitney U test) was used to compare the proportion of cells in each Seurat functional cluster between healthy controls and COVID-19 patients as well as between severe and mild COVID-19 cases. Fisher exact test was used to compare the number of COVID-depleted and COVID-enriched clonotypes being part of each functional Seurat cluster. Multiple testing correction was performed using the Benjamini-Hochberg procedure. Not significant (ns); \* $p < 0.5$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .