Global patterns of antigen receptor repertoire disruption across adaptive immune compartments in COVID-19

Magdalene Joseph\textsuperscript{a,b,1}, Yin Wu\textsuperscript{a,c,d,1}, Richard Dannebaum\textsuperscript{f}, Florian Rubelt\textsuperscript{f,1}, Iva Zlatareva\textsuperscript{ab}, Anna Loren\textsuperscript{c}, Zhipei Gracie Du\textsuperscript{i}, Daniel Davies\textsuperscript{h}, Fernanda Kyle-Cezar\textsuperscript{a}, Abhishek Das\textsuperscript{h,i}, Sarah Gee\textsuperscript{e}, Jeffrey Seow\textsuperscript{c}, Carl Graham\textsuperscript{f}, Dilduz Telman\textsuperscript{f}, Clara Bermejo\textsuperscript{f}, Iva Zlatarev\textsuperscript{a,b}, Anna Loren\textsuperscript{c}, Zhipei Gracie Du\textsuperscript{i}, Thomas S. Hayday\textsuperscript{b}, Isaac Francos-Quijorna\textsuperscript{e}, Shraddha Kamdar\textsuperscript{b}, Richard Davis\textsuperscript{g}, Vasiliki Sofra\textsuperscript{a}, Florencia Cano\textsuperscript{d}, Efstathios Theodoridis\textsuperscript{b}, Lauren Martinez\textsuperscript{d}, Blair Merrick\textsuperscript{c}, Karen Bishauthsingk\textsuperscript{c}, Kate Brooks\textsuperscript{c}, Jonathan Edgeworth\textsuperscript{c}, John Cason\textsuperscript{f}, Christine Mant\textsuperscript{f}, Katie J. Doores\textsuperscript{c}, Pierre Vantourout\textsuperscript{d}, Khi Luong\textsuperscript{f}, Jan Berkal\textsuperscript{d}, and Adrian C. Hayday\textsuperscript{b,2}

Edited by Michael Oldstone, The Scripps Research Institute, La Jolla, CA; received January 28, 2022; accepted May 18, 2022

Whereas pathogen-specific T and B cells are a primary focus of interest during infectious disease, we have used COVID-19 to ask whether their emergence comes at a cost of broader B cell and T cell repertoire disruption. We applied a genomic DNA-based approach to concurrently study the immunoglobulin-heavy (IGH) and T cell receptor (TCR) β and δ chain loci of 95 individuals. Our approach detected anticipated repertoire focusing for the IGH repertoire, including expansions of clusters of related sequences temporally aligned with SARS-CoV-2–specific seroconversion, and enrichment of some shared SARS-CoV-2–associated sequences. No significant age-related or disease severity–related deficiencies were noted for the IGH repertoire. By contrast, whereas focusing occurred at the TCRβ and TCRδ loci, including some TCRβ sequence–sharing, disruptive repertoire narrowing was almost entirely limited to many patients aged older than 50 y. By temporarily reducing T cell diversity and by risking expansions of nonbeneficial T cells, these traits may constitute an age-related risk factor for COVID-19, including a vulnerability to new variants for which T cells may provide key protection.

next-generation sequencing | antigen-receptor repertoires | SARS-CoV-2 | adaptive immune responses | immunoPETE

Severe disease and death caused by SARS-CoV-2 infection appear to be largely due to failures and/or dysregulation of the immune response in vulnerable populations (1). Thus, most SARS-CoV-2 infections are asymptomatic or pauci-symptomatic, particularly among younger people, who, by measures of diversity and functional responses, have greater adaptive immunocompetence than do the elderly (2, 3).

In addition to providing protection, adaptive immune functions may contribute pathologic mediators, including B cell autoreactivities associated with specific disease-related characteristics in many patients with COVID-19 (4, 5). Additionally, early B cell responses after SARS-CoV-2 infection in some donors are enriched in cross-reactive memory B cells, including those against seasonal coronaviruses, which are of uncertain protective benefit (6). Collectively, these examples illustrate the importance of considering the dynamics of immune responses beyond those that are pathogen specific.

Over 2 y since the start of the pandemic, age remains the most evident predisposing factor for COVID-19 severity (7), evoking the increased susceptibility of older persons to other newly emerged viruses, including West Nile virus and SARS-CoV-1 (8, 9). The sequence richness of the CD4\textsuperscript{+} and CD8\textsuperscript{+} T cell receptor (TCR) repertoires is reduced in older persons relative to younger persons (10) and may underlie such vulnerabilities. This raises the question of how age might affect the mobilization of antigen-receptor repertoires in response to SARS-CoV-2 infection.

The COVID-19 pandemic has provided a unique opportunity to interrogate naïve responses to a novel human pathogen. To examine this at scale, we describe here the application of immunoPETE, a newly developed, genomic DNA (gDNA)-based technique permitting simultaneous characterization of antigen-receptor repertoires for the major adaptive lymphocyte subsets: B cells (immunoglobulin heavy chain [IGH] sequences); αβ T cells (TCRβ chain and TRB sequences), which comprise two distinct subsets; CD4\textsuperscript{+} and CD8\textsuperscript{+} cells; and γδ T cells (TCRδ chain and TRD sequences), which comprise Vδ1\textsuperscript{+} and Vδ2\textsuperscript{+} T cells.

Whereas many studies have offered important insights into antigen-receptor repertoires in active COVID-19 (11–15), our study builds on those by uniquely harnessing

**Significance**
While making virus-specific immune responses to SARS-CoV-2, residual B and T cell diversity are key to making responses to concurrent co-infections and/or cancers. We applied an antigen receptor–sequencing technology and provide a comprehensive description of the global immune repertoire in hospitalized and nonhospitalized individuals infected with SARS-CoV-2. Although B cell diversity predictably narrowed, significant narrowing of αβ and γδ T cell diversity was unexpectedly observed only in those aged over 50, which is a major inflexion point for COVID-19–associated mortality. Such a discrepancy in how older persons mount T cell responses to a new virus may be considered a risk factor for the elderly, particularly vis-à-vis new virus variants against which T cell immunity may be particularly important.
an array of technical and methodological approaches. Thus, immunoPETE combines gDNA-based sequencing, used in other COVID-19 repertoire studies (15), with unique molecular identifier (UMI)-based deduplication of clonal expansions and correction of PCR sequencing errors; cell sorting to segregate functionally distinct CD4+ and CD8+ populations; human leukocyte antigen (HLA) typing to enhance clustering and determination of shared specificities within CD4+ and CD8+ T cell populations; and TRD sequencing permitting a real-time comparison of the responses of qualitatively distinct V61 and V62 subsets to a defined human virus infection.

Studying 95 individuals comprising hospital-treated patients with COVID-19 and seropositive [sero(+)] and seronegative [sero(−)] individuals, we found clear adaptive response patterns in each lymphocyte subset studied. Repertoire focusing at the IGH locus was generalizable, with the expansion of related sequence clusters temporally aligned with seroconversion. Whereas TRB repertoires also showed similar shared sequence expansions, these were conspicuously buffered by the scale and diversity of the T lymphocyte compartments in most individuals aged <50 years, whereas they exerted much more disruptive effects in many individuals aged ≥50 years, who likewise displayed major disruptions in TRD repertoire diversity. Such impacts may limit the capacity to recognize diverse challenges (e.g., coinfections and emerging new variants) and may afford undue prominence to expanding nonneutralizing and/or self-reactive clones. Thus, they should be considered an additional age-related disease risk, monitoring of which may help inform the management of COVID-19 and of other infections.

Results

Antigen-Receptor Repertoires in COVID-19. To understand the impact of SARS-CoV-2 infection on T and B cell repertoires, we applied the protocol shown in Fig. 1A to compare 32 hospital-treated patients with COVID-19 (active COVID-19); 20 healthy, convalescent, sero(+) individuals exposed to SARS-CoV-2 who were not hospitalized; and 43 nonexposed, healthy, sero(−) adult control individuals. For all participants (SI Appendix, Table S1), detailed immune profiling, serological analyses, and detailed clinical annotation (including World Health Organization severity scoring) were available based on their

**Not all healthy control samples were run through the full COVID-IP Study pipeline. All samples had SARS-CoV-2 serology data. (B) Recovery of CDR3s from sero(−) (n = 47), sero(+) (n = 26), mild (n = 10), moderate (n = 26), and severe (n = 16 CD4+; n = 15 CD4−) fraction samples by immunoPETE. gDNA (2 × 250-ng replicates) was used for CD4+ library preparation (prep) and 1,000 ng of gDNA (4 × 250-ng replicates) was used for CD4− library preparation. Bar indicates the median value. Mann-Whitney U test results are shown. (C) Correlation between immunoPETE and flow cytometry (cyt). Percentage of TRB, IGH, and TRD CDR3s recovered by immunoPETE sequencing in the CD4+ fraction (y-axis) versus percentage of CDR3 (n = 112), CD8 (n = 112), and TRC (n = 112) assayed by flow cytometry (x-axis) of the same sample. Spearman’s correlation was used for analysis. (D) Correlation of overall CD4+ TRB (Left) and CD4− TRB (Right) diversity with frequency (fraction) of naïve (CD45RA−CCR7+) CD4+ (n = 107) or CD8+ T cells (n = 106), respectively, as previously reported in the COVID-IP Study. Spearman’s correlation was used for analysis. Note that the color code for each cohort is maintained, and unadjusted P values are displayed. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

Fig. 1. Application of immunoPETE to the COVID-19 Study cohort enables efficient, high fidelity, and quantitative recovery of TRB, IGH, and TRD CDR3s from peripheral blood mononuclear cells (PBMCs). (A) Summary of workflow for samples recruited into the present study. MACS: magnetic activated cell sorting. *Not all donors had longitudinal blood sampling.
enrollment in the COVID immunophenotyping (COVID-IP) study (16).

Many individuals were aged ≥50 y (SI Appendix, Fig. S1A and Table S1), and because this is reportedly the most overt infection point in the risk of COVID-19–associated death (17–19), we used age to segregate the antigen-receptor repertoire analyses, which were also segregated into those obtained up to 14 d after symptom onset (“early”), those obtained after that time (“late”), and those who were either asymptomatic sero(+) individuals in the community or asymptomatic hospitalized donors admitted for unrelated conditions (“unknown”). This temporal division is important for patients positive for COVID-19 who were aged ≥50 y to be compared with younger patients who invariably recovered earlier. Note that the color code used for each cohort in Fig. 1 is maintained throughout all data figures. As expected, the severe COVID-19 subcohort was enriched in male patients (SI Appendix, Fig. S1B).

gDNA-Based Repertoire Sequencing. First, CD4+ T cells were purified from freshly harvested peripheral blood mononuclear cells by magnetic bead separation (Fig. 1A). We considered the CD4-depleted (CD4+) T cell fraction as primarily a de facto source of CD8+ αβ T cells (since blood harbors very few CD4−CD8− αβ T cells); B cells; and γδ T cells.

We selected the IGH, TRB, and TRD loci for global assessment, because they encode greater reservoirs of repertoire diversity than do the immunoglobulin (Ig) light chain, TCRα, and TCRγ loci, which also have leuker allic exclusion that confounds correspondence of productively rearranged gene sequences with cell numbers. Furthermore, TCRγ rearrangements are common in αβ T cells and prevent use of the γ chain to assess γδ T cell diversity in unsorted populations (20).

Antigen-receptor repertoires were assessed using immunoPETE, an assay that includes primer mix targeting all known human IGHV, IGJH, TRBV, TRBJ, TRD and TRDJ genes, and provides unprecedented capability to concurrently sequence all three antigen-receptor chains from one sample. This is of particular importance in lymphopenic settings such as COVID-19 where clinical material may be very limited. The use of UMs, which have hitherto mostly been applied to RNA-based antigen-receptor repertoire analyses, allows the numbers of individual clonal cells with an identical V-CDR3-J sequence to be quantified with greater confidence. Moreover, it allows identification of sequencing and PCR errors, which are apparent as rare variants of a consensus sequence with identical UMs (SI Appendix, Fig. S2 A and B). The power of this error suppression is evidenced by the application of immunoPETE to two clonal T cell lines, HuT78 and Jurkat, with >99% of gDNA templates recovered being of a monoclonal origin for both cell lines (SI Appendix, Table S2).

Template lengths range from 170 to 210 nucleotides (interquartile range), thereby capturing comprehensive information on V, D, and J gene segment usage and on unique P and N nucleotide–mediated contributions to each CDR3. All nonproductive V-D-J rearrangements were excluded, as were any artificial hybrid sequences. Real-world assay reproducibility was evidenced when we assessed technical quadruplicate libraries from the CD4+ fraction, using 250 ng of input DNA per library and quadruplicate libraries from the CD4+ fraction, which contains multiple cell types at lower representation. In each case, there was extremely high concordance between replicates in the percentage of each chain recovered, with mean SEs between 0.05% and 0.49% (SI Appendix, Fig. S3). Additionally, immunoPETE could reproducibly determine repertoire diversity, measured by Shannon entropy (a.k.a., Rényi α = 1) for which normalized values approaching 1.0 reflect diverse, polyclonal repertoires (21). Again, use of replicates showed very high concordance with mean SEs in Shannon entropy, ranging from 0.0004 to 0.0093 (SI Appendix, Fig. S4).

Any sequencing approach is theoretically vulnerable to amplification biases introduced during multiplex PCR, attributable to, for example, mispriming or selectively lesser priming efficiency. However, all target TRBV genes were recovered with relatively similar hierarchies in the CD4+ and CD4− fractions (SI Appendix, Fig. S5 A and B) and, overall, there was good correlation for the relative representations of TRBV genes in both CD4+ and CD4− fractions across donors (SI Appendix, Fig. S5 C and D), albeit less so in the CD4− fraction, consistent with the greater oligoclonality (reduced entropy) of CD8+ versus CD4+ T cells (SI Appendix, Fig. S4). These data demonstrate the consistent performance of immunoPETE such that any primer bias does not preclude intrastudy comparisons. Nonetheless, to permit interstudy comparisons, we sought to assert the comparability of immunoPETE outputs with other independent, publicly available datasets. Indeed, the TRBV and TRBJ hierarchies identified by immunoPETE correlated significantly with published datasets of TRBV and TRBJ gene usage in adult peripheral blood determined either by other sequencing methods (22, 23) or by flow cytometry (24) (SI Appendix, Fig. S6 A and B). Thus, immunoPETE consistently recovered all TRBV genes and displayed no gross amplification biases when benchmarked against independent assays.

The effective practical application of immunoPETE was evident in our obtaining on average ~23,000 productive TRB CDR3s from 500 ng (2 × 250-ng technical replicates) of DNA derived from the CD4+ fraction (Fig. 1B). The slightly lower recovery of CD4+ TRB CDR3s from patients with active COVID-19 likely reflects COVID-19–associated T cytopenia that was particularly severe for CD8+ T cells and VD2+ T cells (16, 25–27) and that was reflected in TRB and TRD showing the poorest sequence recoveries from the CD4− fraction (Fig. 1B).

Importantly, there was a striking correlation of recovered IGH, TRB, and TRD CDR3 sequences with flow cytometric enumeration of cells in aliquots parallel to the sequenced samples (Spearman’s r > 0.8; P < 0.0001) (Fig. 1C). Given this robust and quantitative nature of immunoPETE, we could draw equivalence of productively rearranged CDR3s to cell counts, a frame of reference used from this point on. In total, 250 aggregate next-generation sequencing libraries, one each for CD4+ and CD4− fractions, were constructed and analyzed, including some longitudinal samples primarily from patients with active COVID-19. Our analysis provides data derived from ~6.5 million antigen-receptor chain sequences. In a rapid and basic biological test of this dataset, we observed highly significant correlations of Shannon entropy with the frequencies of CD45RA+CCR7+ αβ T cells that largely denote unexpanded, naïve T cells that would be expected to be highly diverse (Fig. 1D). Likewise, we found strong and inverse correlations of Shannon entropy with the frequencies of canonical memory T cell subsets that would be expected to be clonally expanded (Fig. 1E). Because different measures of population diversity can be variably affected by, for example, dominant clone sizes, we supplemented Shannon entropy measures with measures of D50 (the proportion of unique clones, ordered by dominance, accounting for 50% of all CDR3s; low values reflect clonal expansions), and of Simpson’s dominance (high values reflect clonal expansions). Again, we observed consistent, strong, and significant correlations of D50 and Simpson’s dominance with the frequencies of canonical memory T cell subsets (Fig. 1E).
IGH Sequence Focusing and Clustering in COVID-19. Assessments of individual protection against SARS-CoV-2 are mostly based on antibody titers. Thus, with immunoPETE established as a robust method for tracking lymphocyte repertoires, we first interrogated the effect of COVID-19 on the IGH repertoire. Indeed, Shannon entropy and Simpson’s dominance measurements revealed significant IGH focusing (i.e., overrepresentation of IGH sequences suggestive of clonal expansion) for individuals sampled within 14 d of symptoms relative to sero(-) control participants (Fig. 2A). By D50, there was significant focusing in those aged <50 y and a trend for those aged ≥50 y (Fig. 2A). Donors with significant focusing spanned a spectrum of disease severities, ages, and clinical features. However, focusing had largely renormalized at later time points, suggesting a strong temporal association with disease duration (Fig. 2A).

It might be expected that sequence focusing following virus exposure would be reflected in expansions of structurally similar clones, with shared specificities. Using the “DefineClones.py” script in the “change-o” toolbox (SI Appendix, Materials and Methods), we found that, relative to samples from age-matched sero(-) individuals, early samples from SARS-CoV-2-exposed individuals aged <50 y showed a significant increase in the

**Fig. 2.** IGH clustering and repertoire focusing associated with SARS-CoV-2 exposure. (A) Overall IGH repertoire diversity assessed by Shannon entropy, Simpson’s dominance (S dom) and D50. The bar indicates the median value. Kruskal-Wallis test with post hoc Dunn’s test against age-matched sero(-) control individuals was used for analysis; the unadjusted P values are shown. (B) IGH CDR3s clustered (clust) as a proportion of total IGH CDR3s. The bar indicates the median. Kruskal-Wallis test with post hoc Dunn’s test against age-matched sero(-) control individuals was used for analysis; unadjusted P values are shown. (C) Within cluster IGH repertoire diversity assessed by Shannon entropy, D50, and Simpson’s dominance. The bar indicates the median. Kruskal-Wallis test with post hoc Dunn’s test against age-matched sero(-) control individuals was used for analysis; the unadjusted P values are shown, (D) IGH repertoire focusing within clusters assessed by Shannon entropy (Left) as well as normalized (norm) anti-spike IgM titers (Right) plotted by time from symptom onset (1-wk bins). (E) Heat map of correlations with IGH Shannon entropy in individuals with active COVID-19. Color scale denotes Spearman’s r with corresponding P values indicated within each cell. Nonsignificant correlations (P > 0.05) were left blank. (F) Fraction of IGH CDR3s occupied by the sum of CoV-AbDab IGH CDR3 matches per sample (including clustered-related sequences). The bar indicates the median. Kruskal-Wallis test with post hoc Dunn’s test against age matched sero(-) control individuals was used for analysis; the unadjusted P values are shown. (G) Average representation of CoV-AbDab IGH CDR3 matches measured by fraction of total IGH CDR3s per sample (including clustered-related sequences). The bar indicates the median. Kruskal-Wallis test with post hoc Dunn’s test against age matched sero(-) control individuals was used for analysis; the unadjusted P values are shown. (H) Heat map of correlations with the total repertoire fraction occupied by CoV-Ab Dab IGH CDR3 matches (including related clustered sequences). Color scale denotes Spearman’s r with corresponding P values indicated within each cell. Nonsignificant correlations (P > 0.05) were left blank. (A–C) Age: <50 y: sero(-) (n = 30); <50 y: early (n = 9); <50 y: unknown (n = 23); ≥50 y: sero(-) (n = 17); ≥50 y: early (n = 17); ≥50 y: late (n = 17); and ≥50 y: unknown (n = 11). (B) IGH Shannon entropy; sero(-) (n = 47); 1 wk (n = 14); 2 wk (n = 12); 3 wk (n = 7); 4 wk (n = 4); 5 wk (n = 4); and ≥6 wk (n = 3). For spike IgM: sero(-) (n = 34); 1 wk (n = 14); 2 wk (n = 12); 3 wk (n = 7); 4 wk (n = 5); 5 wk (n = 2); and ≥6 wk (n = 3). Freq, frequency; Unk, unknown.
fraction of total IGH sequences that could be found in structurally related clusters. This was also true as a trend for those aged ≥50 y, and it was overt in several participants across a range of disease severities (Fig. 2B). The clustered fraction of the repertoire has a significantly lower Shannon entropy in SARS-CoV-2–exposed individuals than in sero(−) control participants, diagnostic of significant clonal expansions within the clusters (Fig. 2C). Additional metrics also revealed significantly reduced diversity of clustered IGH sequences from SARS-CoV-2–exposed individuals (Fig. 2C). We found a temporal drop in Shannon entropy over a time frame of ~3 wk after symptom onset, slightly preceding but largely coincident with the development of spike-specific IgM (Fig. 2D).

**IGH Repertoire Architecture and Antigen-Specific Responses.**

We next found that IGH focusing (low Shannon entropy) in individuals with active COVID-19 strongly correlated with high plasmablast frequency measured by flow cytometry and, albeit more modestly, with IgM antibodies to spike, RBD, and N proteins (Fig. 2E). To further assess whether the observed focusing was likely driven by SARS-CoV-2–reactive clones, we retrieved IGHV-CDR3-IGHJ sequences from the CoV-AbDab database (28), which contained 2,203 fully human SARS-CoV-2 reactive IgS as of July 9, 2021. Although antigen specificity reflects VH and VL sequence composition, VH chains commonly make dominant contributions (29). We found 51 instances of exact IGHV-CDR3-IGHJ matches, representing 23 unique sequences, some of which were found in multiple samples. Moreover, almost half (n = 11 of 23) (SI Appendix, Table S3) were contained within the aforementioned IGH clusters, which we showed to be expanded in SARS-CoV-2–exposed donors (Fig. 2C). When we included IGHV-CDR3-IGHJ sequences that were very close relatives of the direct matches, a total of 33 CoV-AbDab–associated sequences matches (n = 61) were found across multiple samples.

As anticipated, CoV-AbDab–associated sequences were significantly enriched in SARS-CoV-2–exposed individuals (active COVID-19, n = 21; sero(+), n = 6) relative to sero(−) individuals (n = 3; P < 0.0001 by Fisher’s exact test) (SI Appendix, Table S4). Moreover, those sequences were expanded within individuals with active COVID-19, relative to sero(−) individuals, since they collectively, and on average, accounted for a significantly higher fraction of an individual’s total repertoire (Fig. 2 F and G). In individuals harboring such sequences, the fraction of the repertoire to which they contributed, albeit small, was strongly correlated with repertoire focusing, with SARS-CoV-2 serology, and with plasmablast frequencies (Fig. 2H) consistent with SARS-CoV-2 reactive clones driving repertoire focusing, as would probably have been expected. Nonetheless, the benefit of SARS-CoV-2 reactive antibodies cannot be universally assumed. In fact, the most widely shared IGH sequence between our cohort and CoV-AbDab (CDR3-CARGFDYW) was found exclusively in SARS-CoV-2–exposed individuals (n = 17 of 52) but contributes to nonneutralizing S-reactive antibodies (28, 30, 31) and has been reported in Kawasaki disease (32) (see Discussion).

**Changes in TRB Repertoires in COVID-19.**

We next turned our attention to the impact of COVID-19 on αβ T cell repertoires from the same individuals for whom IGH data had been obtained. First, we asked whether there was any evidence for viral superantigen-mediated polyclonal targeting of T cells expressing defined Vβ genes, since this has been considered to occur in COVID-19 (33) and in post–SARS-CoV-2–associated, pediatric, multisystem inflammatory syndrome (34). Our analyses failed to find significant differences between individuals with active COVID-19 and sero(−) individuals in the representation of any of TRBV5-6, TRBV11-2, TRBV13, TRBV14, or TRBV24-1 genes that were collectively implicated in those studies (Fig. 3A and SI Appendix, Fig. S7).

Nonetheless, there were modest but significant differences in the representation of a small number of V genes among CD4+ and CD8+ T cells of individuals with active COVID-19 versus sero(−) control individuals (Fig. 3A and SI Appendix, Fig. S7). Strikingly, TRBV6-4, which showed the most significant difference in the CD8+ compartment, being underrepresented in the active COVID-19 cohort, is disproportionately expressed by “unconventional” CD8+ mucosal-associated invariant T cells that have been described as severely cytopenic in COVID-19 (35) (Fig. 3A). In sum, COVID-19 was associated with some significant Vβ-associated changes, but the overall architecture of the αβ TCR repertoire remained diverse and lacked any dominant TRBV gene representation that would reflect overt superantigen responses.

**Age-Related TCRVβ Focusing in COVID-19.**

As for IGH, we next investigated how the overall architecture of the TRB repertoire was affected by COVID-19. Surprisingly, we found that by contrast to IGH, significant reductions in Shannon entropy and increases in Simpson’s dominance were not a general trait of the CD4+ compartment, but rather were limited to SARS-CoV-2–exposed individuals aged ≥50 y sampled within 14 d of symptom onset (Fig. 3B). Compared with CD4+ TRB, entropy values for CD8+ TRB were lower among sero(−) individuals (Fig. 3C), consistent with the greater degree to which CD8+ T cells show clonal expansions in response to myriad environmental exposures (36, 37). Nonetheless, the CD8+ TRB repertoires of SARS-CoV-2–exposed individuals aged ≥50 y sampled within 14 d of symptom onset showed a significant increase in Simpson’s dominance and a trend toward decreased Shannon entropy and D50 compared with sero(−) control individuals (Fig. 3C). Again, this was not a collective trait of those aged ≥50 y, as sero(−) and sero(+) (“unknown”) individuals of both age groups showed comparable diversity metrics. Thus, the differences seen in SARS-CoV-2–exposed persons ≥50 y are a function of age and of disease rather than of either alone. Suggesting similar underlying mechanisms in both αβ compartments, the degree of CD4+ TRBV–repertoire focusing correlated strongly with CD8+ TRBV–repertoire focusing. Additionally, it is evident from the color-coded data (Fig. 3 B and C) that those with the highest degree of focusing spanned a spectrum of disease severities.

To ensure that these conspicuous, age-related impacts on the αβ T cell compartment did not simply reflect lower numbers of sequences recovered (a potentially confounding variable in lymphopenia), each sample was subsampled to 1,200 (CD8+) or 2,400 (CD4+) cells (samples with fewer than these numbers of cells were excluded), and medians of metrics that were computed from 100 such resamples were reported as “subsampled” values. Strikingly, reductions in entropy and increases in dominance were now even more overt for those aged ≥50 y, being significant for CD4+ and CD8+ cells by all measures (SI Appendix, Fig. S8 A and B), whereas no significant changes were observed for patients with COVID-19 aged <50 y (SI Appendix, Fig. S8 A and B). Interestingly, by every entropy or dominance metric, the TRB repertoires of those aged ≥50 y returned toward normal at later time points after symptom onset, indicating a potential for renormalization.
Fig. 3. Age-related repertoire changes in CD4+ and CD8+ T cells associated with COVID-19. (A) Box plots of CD4+ TRBV V gene family use as a proportion of the unique CDR3s (i.e., each unique CDR3 is treated equally regardless of clone size) in samples (n = 47) from sero(-) individuals and samples (n = 51) from individuals with active COVID-19. TRBV genes with significantly different utilization between cohorts are highlighted in red. Median, interquartile range, and range values are plotted. The Mann–Whitney U test was used for analysis; unadjusted P values are displayed. *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001. (B) Overall CD4+ TRB repertoire diversity assessed by Shannon entropy, D50, and Simpson’s dominance (S dom). For CD4+ TRB metrics: age <50 y: sero(-) (n = 30); <50 y: early (n = 9); <50 y: unknown (n = 23); ≥50 y: sero(-) (n = 17); ≥50 y: early (n = 17); ≥50 y: late (n = 18); and ≥50 y: unknown (n = 11). (C) Overall CD4+ TRB repertoire diversity assessed by Shannon entropy, D50 and Simpson’s dominance. For CD4+ TRB metrics: age <50 y: sero(-) (n = 30); <50 y: early (n = 9); <50 y: unknown (n = 23); ≥50 y: sero(-) (n = 17); ≥50 y: early (n = 17); ≥50 y: late (n = 18); and ≥50 y: unknown (n = 11). The bar indicates the median. Kruskal-Wallis test with post hoc Dunn’s test against age-matched sero(-) control individuals; unadjusted P values are shown. (D) Defining SARS-CoV-2–associated TCR clusters. Circles represent individuals (identified by numbers) and exposure to SARS-CoV-2 (color). Clusters are built from all TCRs present in individuals with a specific HLA gene, in this example, HLA B*07:02, by grouping similar CDR3s. We tested each of such raw clusters for overrepresentation of SARS-CoV-2–exposed (active COVID-19 and sero(+)) or sero(-) individuals by Fisher’s exact test (significance set at P < 0.05). A representative cluster built from CD4+ TRB sequences of one sero(-) individual and six SARS-CoV-2–exposed individuals on an HLA B*07:02 background from the study by Shomuradova et al. (17). Overlap was present in two of eight individuals with active COVID-19, 4 of 9 sero(+) individuals and 0 of 17 sero(-) individuals (indicated by the shaded cell). Color scale indicates the proportion of total CD4+ repertoire in each sample occupied by overlapped clones.
Given that CD8\(^+\) T cells are documented to be more clonally focused than CD4\(^+\) T cells, there was a theoretical potential for our observations to be confounded by the purity of CD4\(^+\) and CD8\(^+\) fractions, respectively. Therefore, we examined the relationship between fraction purity and TRB clonal metrics. Evidently, the positively selected CD4\(^+\) fraction was very pure, with all but six samples containing <1% contamination by CD8\(^+\) T cells. Unsurprisingly, there was a correlation of clonal focusing with the degree of CD4\(^+\) contamination, but it was weak and clearly did not drive the marked clonal focusing we observed in individuals aged $\geq 50$ y (SI Appendix, Fig. S9A). Indeed, exclusion of the six outliers with >1% contamination by CD4\(^+\) T cells did alter the pattern of our observations for the CD4\(^+\) fraction (SI Appendix, Fig. S9B). Predictably, the CD4\(^+\) fraction was enriched for CD8\(^+\) T cells but had a variable degree of contamination by CD8\(^-\) (CD4\(^+\) and double-negative) T cells. Importantly, we observed no correlation in the CD4\(^-\) fraction between purity and clonal metrics (SI Appendix, Fig. S9C).

**TRBV Sequence Clustering in COVID-19.** We next sought evidence for repertoire changes imposed by antigen-specific responses to SARS-CoV-2. Because the TCRs of conventional $\alpha\beta$ T cells are major histocompatibility complex (MHC) restricted, we first used DNA typing to identify discrete MHC class I ($n = 21$) and class II alleles ($n = 28$) shared across at least eight individuals, and for each of those HLA alleles, we then assessed TRBV-CDR3-J sequences ($>2$ million unique CD4\(^+\) TRB sequences; ~900,000 unique CD8\(^+\) TRB sequences) for clustering (SI Appendix, Table S5). As an example, 43,957 CD8\(^+\) TRB sequences were contributed by 9 HLA-B*07:02\(^+\) SARS-CoV-2–exposed individuals, while 81,342 CD8\(^+\) TRB sequences were contributed by 11 HLA-B*07:02\(^-\) sero(−) individuals (Fig. 3D).

Given the diversity of HLA alleles of individuals in our study, it was not possible to match all alleles for each clustering run. Thus, we could not unequivocally attribute all of the aforementioned CD8\(^+\) TRB sequences to HLA-B*07:02–restricted T cells, although it seems logical to assume that the sequences will be highly enriched in those contributing to HLA-B*07:02–restricted TCRs. Therefore, we proceeded to cluster the sequences using GLIPH2 (38), which infers similar peptide-MHC specificity on TCRs. Therefore, we proceeded to cluster the sequences using GLIPH2 (38), which infers similar peptide-MHC specificity on TCRs. Thus, we could not unequivocally attribute all of the aforementioned CD8\(^+\) TRB sequences to HLA-B*07:02–restricted T cells, although it seems logical to assume that the sequences will be highly enriched in those contributing to HLA-B*07:02–restricted TCRs. Therefore, we proceeded to cluster the sequences using GLIPH2 (38), which infers similar peptide-MHC specificity on TCRs.
CD45RA⁺CD27⁻ V61⁺ T cells (Fig. 4D), which emerged as one of two only immunological correlates of virus titers in the COVID-IP study, the other being natural killer cell numbers (16). Clonal expansions of human CD45RA⁺CD27⁻ V61⁺ cells have been observed in the blood in several settings, including cytomegalovirus (CMV) reactivation (41–44), but this is seemingly the first clear example of such an adaptive response in relation to acute viral infection.

There was a statistically significant correlation of TRDV1 focusing with CD4⁺ and CD8⁺ TRB focusing (Fig. 4B), but in contrast to TRB or IGH sequences (Fig. 3E and SI Appendix, Table S3), very few TRDV1 sequences were shared across donors, and none was significantly enriched in SARS-CoV-2–exposed individuals, suggesting that the V61 responses in COVID-19 are almost exclusively private. Likewise, very few public sequences were reported in other settings of V61 focusing (42). Likely, V81⁺ TCRs are mostly not specific for pathogen-derived antigens but react to myriad molecular sentinels of dysregulation directly resulting from virus infection. Indeed, the CDR3 lengths, which can be highly variable for TRDV1, showed comparable distributions across SARS-CoV-2–exposed individuals, sero(−) individuals, and a prior analysis of human breast-resident V61 TCRs (45) (SI Appendix, Fig. S11).

Age-Related V62 Cell Losses. Consistent with the prominence of V62⁺ T cells in peripheral blood, most TRD CDR3s were TRDV2-D-J rearrangements in sero(−) individuals of all ages tested. This was also true for SARS-CoV-2–exposed individuals aged <50 y, albeit there was interindividual variation as is well established for peripheral blood Vγ9V82⁺ cells (46, 47)
Fig. 5. Age-related selective depletion of Vβ2 T cells in SARS-CoV-2-exposed donors. (A) Vβ2 T cells (TRDV2) as a percentage of total γδ T cells (TRD). (B) d97-LVI PAg-reactive TRDV2 sequences recovered as a percentage of total TRDV2 sequences recovered. (A and B) Age: <50 y: sero(-) (n = 30); <50 y: early (n = 9); <50 y: unknown (n = 23); ≥50 y: sero(-) (n = 17); ≥50 y: early (n = 16); ≥50 y: late (n = 17); and ≥50 y: unknown (n = 11). (C) d97-LVI PAg-reactive TRDV2 sequences recovered as a percentage of total TRDV2 sequences recovered plotted by severity of COVID-19 disease [sero(-), n = 17; sero(+), mild, n = 11; moderate-severe, n = 33]. (D) Overall TRDV2 repertoire diversity assessed by Shannon entropy. Only samples with ≥30 TRDV2 sequences were analyzed for repertoire diversity and plotted. The bar indicates the median. Kruskal-Wallis test with post hoc Dunn’s test against age-matched sero(-) control individuals was used for analysis; the unadjusted P values are shown. Age: <50 y: sero(-) (n = 30); <50 y: early (n = 9); <50 y: unknown (n = 23); ≥50 y: sero(-) (n = 17); ≥50 y: early (n = 13); ≥50 y: late (n = 16); and ≥50 y: unknown (n = 9).

Discussion

In this study, we have applied a protocol for gDNA-based antigen receptor sequencing to assess at-scale the concurrent impacts of infection on the IGH, TRB, and TRD repertoires. Being gDNA based, our approach provided information about cell numbers that strongly correlated with flow cytometry data and may, therefore, prove powerful for samples such as post-mortem formalin-fixed, paraffin-embedded tissues, that are refractory to other forms of cell enumeration. Erroneous sequence calls were greatly reduced by including UMI, which is particularly important in relation to template-independent CDR3 sequences and somatic mutation in Ig genes. Aside from vaccination, there have been few opportunities to measure the establishment of adaptive immune responses to a known challenge incurred at a reasonably well-defined time point. Indeed, there have been few opportunities to measure the architectures of such responses in the context of overt immune dysregulation, sometimes in life-threatening settings, as was the case for some of the patients in the COVID-IP study from whom we obtained antigen-receptor sequences (16). It is, therefore, noteworthy and encouraging that despite such multifaceted immune dysregulation, extensive and dynamic sequence clustering in the IGH and TRB repertoires, together with some overlap with SARS-CoV-2-associated sequences and some level of sequence sharing, albeit minor, was strongly suggestive of virus antigen–specific responses across age groups and across disease severity. Thus, immunoPETE has demonstrated robust characterization of antigen repertoires in real-world clinical samples.

In the face of the continuing COVID-19 pandemic, vaccination and serological monitoring are backbones of the public health response, in which regard it is noteworthy that B cell repertoire dynamics were largely comparable across age groups and severity indices. By contrast, the T cell compartments...
Disruptive T cell repertoire responses in SARS-CoV-2 are associated with age. Principal component analyses (PCAs) of age-related T cell repertoire responses (components included: CD4⁺ TRBV, CD4⁺ TRBV, and TRDV1 Shannon entropy, D50, and Simpson’s dominance, and TRDV2 % of TRD and TRDV2 δ% LVI percentage) in sero(-) samples and SARS-CoV-2-exposed samples (restricted to the “early” cohort to avoid bias from time from symptom onset). (A) PCA of T cell repertoire responses in samples from donors aged ≥50 y (Left) colored by severity, and box plot of principle component (PC) 1 split by SARS-CoV-2 exposure (Right), demonstrating a significant difference in PC1 between cohorts. (B) PCA of T cell repertoire responses in samples from donors aged <50 y (Left) colored by severity, and box plot of PC1 split by SARS-CoV-2 exposure (Right), demonstrating no significant difference in PC1 between cohorts. (C) PCA of T cell repertoire responses in samples from donors aged ≥50 y (Left) shaded by sex, and box plot of PC1 split by sex (Right), demonstrating no significant difference in sex composition. (D) PCA of T cell repertoire responses in samples from donors aged <50 y (Left) shaded by sex, and box plot of PC1 split by sex (Right). Mann-Whitney U test was used for analysis.

(i.e., αβ and γδ) unexpectedly diverged from the B cell compartment in demonstrating markedly different repertoire responses to COVID-19 in those aged ≥50 y versus those aged <50 y. This evokes flow cytometric and functional evidence for age-related chaotic dysregulation in the adaptive immune compartment in COVID-19 (50). Unfortunately, a limitation of this study was the rarity of longitudinal samples obtained at equivalent sampling intervals or at equivalent points after symptom onset. Thus, there was insufficient power to detect any clear trajectories of high-abundance clones collectively composing the top 1%.

Given that age remains one of the most significant risk factors for infectious-disease severity, even after correction for age-related comorbid conditions (7), one may consider whether age-associated exaggerations of TRB and/or TRD repertoire focusing may be contributory factors. By definition, exaggerated focusing would limit T cell diversity that may be particularly important vis-à-vis antibody-escape variants of concern and/or other coinfections. Additionally, exaggerated focusing may include disproportionate bystander expansions with potential to be immunopathogenic. In COVID-19, these TCR-repertoire disruptions were also set against a backdrop of marked age-related losses of V62 sequences required to maintain innate-like responsiveness to PAgS that are a sentinel of infection by myriad bacteria and viruses. Collectively, these points are particularly germane to a scenario by which COVID-19 is becoming endemic and, as the world emerges from a zero-COVID policy, the burden of morbidity will fall disproportionately on the elderly.

A largely unresolved immunological question of clinical importance is whether adaptive responses to virus infection are limited to antigen-specific B and αβ T cells, or whether they include potentially protective responses of human γδ T cells, for which the adaptive versus innate status has been oft debated (51–53). While V61⁺ clonal expansions were reported in human blood and liver (42, 43, 54), the provoking stimuli are largely unelucidated, with data supporting and questioning, respectively, a role for CMV (42, 43). By contrast, our study has provided an unique association of bona fide adaptive V81 responses with a live virus challenge, as evidenced by a correlation of decreased TRDV1 entropy with V81⁺ cell expansions that, in turn, correlated with virus load (16). However, the lack of any COVID-19–associated sequence sharing supports, prima facie, the prospect that expanded γδ cell sequences may reflect reactivity to virus-induced changes in endogenous antigen expression rather than to virus (55). This underscores the importance of assessing immunological responses beyond those that are directly SARS-CoV-2 specific.

In considering the causes of age-related T cell repertoire disruptions, it is probably inappropriate to consider that those
aged ≥50 y are intrinsically T cell immunodeficient since they ordinarily harbor rich CD4+ and CD8+ T cell repertoires (10). However, those aged ≥50 y may harbor greater percentages of senescence-associated CD57+ CD28− p16+ T cells that are refractory to clonal expansion (56), perforce leaving responses to be dominated by larger expansions of smaller numbers of clones. Moreover, an identical outcome would result from the need to expand antigen-specific, naïve T cells in order to combat newly emerging pathogens, since the naïve T cell compartment in older persons has a very uneven geometry featuring many distinct, highly unequal expansions of private specificities (10).

An alternative, but not mutually exclusive, viewpoint is that marked clonal focusing at early time points in individuals aged ≥50 y may reflect a prompt recruitment and expansion of pre-existing T cells primed to related antigens (e.g., seasonal coronaviruses) (57, 58), which those aged ≥50 y are more likely to harbor. However, whether those T cells are immunoprotective against SARS-CoV-2 remains unclear (57, 58). This evokes similar findings demonstrating that early B cell responses after SARS-CoV-2 infection were enriched in cross-reactive memory B cells, including against seasonal coronaviruses, that were of uncertain protective benefit, whereas late B cell responses were enriched in neutralizing SARS-CoV-2 RBD-specific B cells (59).

The disruptive response architectures of those aged ≥50 y would not necessitate a poor outcome; they would simply constitute a potentially significant added risk factor contributing to the upward inflection point for the probability of COVID-19–associated death for those aged ≥50 y. Note that our attempts to parse responses into other age groups failed to identify a clearer segregation of TCR responses than those between people older or younger than 50 y, respectively. Such disruptive adaptive immunity might likewise have contributed to the added risk incurred by older persons encountering SARS-CoV-1 and West Nile virus, when those infections first emerged in humans (8, 9). And it may likewise contribute to relatively poor first-dose immunogenicity in older persons and patients with cancer of vaccines against neo-antigens, including SARS-CoV-2, whereas memory T cell responses to CMV, Epstein-Barr virus, influenza, and tetanus appeared largely normal (60–62).

In sum, there are many potential hazards of reduced TCR diversity that are a collateral cost of mobilizing SARS-CoV-2/COVID-19–reactive T cell immunity in many persons aged ≥50 y. As such, this is an age-related risk factor for which clinical decision-making might be enhanced by readily available information on the status of the TRB and TRD repertoires in COVID-19 and in other infectious diseases.

Materials and Methods
Study Design and Participants. Peripheral blood samples were obtained from 95 donors, including patients treated for COVID-19 at Guy’s and St. Thomas’ NHS Trust (London, UK); and both sero(−) and sero(+) healthy control individuals based at King’s College London (full cohort details are listed in SI Appendix, Table S1). All donors were recruited as part of the COVID-19 study (16) and had given written informed consent. This study was approved by the ethics approval board of the King’s College London Infectious Diseases Biobank (REC 19/30/ O223). Paired serology, flow cytometry, cytokine assay, and SARS-CoV-2 PCR data were available for most donors through their enrollment in the COVID-19 study. Full details of study design and donor cohorts are provided in the SI Appendix, Materials and Methods.

Magnetic Bead Separation and Generation of Antigen-Receptor Libraries Using ImmunoPETE Protocol. Peripheral blood mononuclear cells isolated by Ficoll gradient separation were split into CD4+ and CD8− fractions using magnetic bead separation (Miltenyi, CD4 Microbeads). Genomic DNA was isolated from both fractions and used to generate quantitative TCRβ, TCRδ, and immunoglobulin-heavy chain libraries from a single-well assay using the immunoPETE protocol from Roche Sequencing Solutions. The immunoPETE protocol combines V gene priming and extension with UMI-tagged oligos, followed by J primer extension and library amplification of bead-purified PCR products from the initial V primer extension. The full experimental protocol is provided in the SI Appendix, Materials and Methods.

Data Processing and Statistical Analysis. A Roche in-house bioinformatics pipeline was used to process sequencing data with additional use of R, version 4.0.3, CRAN available packages. Prism 9 (GraphPad Software) was used for data visualization. Statistical tests were performed using either Prism 9 or R, version 4.0.3, and are specified in accompanying figure legends. All tests were two-sided unless otherwise specified. Full details of data processing and analysis are provided in the SI Appendix, Materials and Methods.

Data Availability. Sequencing data is available in the AIRR Data Commons (63) and is accessible using the iReceptor Gateway (http://gateway.ireceptor.org/) (64) with study accession ID IR-Roche-000001. Software for immunoPETE is available at https://github.com/bioinform/Daedalus (65).

Acknowledgments. We thank all patients and blood donors who participated in this study, as well as the medical and research teams at Guy’s and St. Thomas’ Hospitals. Schema figures were created by BioRender.com. We likewise thank our colleagues for discussions and for their support. This study was supported by a Cancer Research Institute Irvington Fellowship (D.R.M.); Cancer Research UK (CRUK) Cancer ImmunoTherapy Accelerator (A.C.H., I.D.M.B.D.); CRUK City of London Major Cancer Centre studentship C640/A27440 (M.J.); European Molecular Biology Organization Long-term Fellowship 198-2018 (M.M.R.); Francis Crick Institute, which receives core funding from CRUK (grant FC001093), Medical Research Council (MRC) grant FC001093, and the Wellcome Trust (grant FC001093; A.C.H., L. Monin, and F.C.); John Black Charitable Foundation (A.C.H.); King’s Together Rapid COVID-19 Call Award (K.J.D.); King’s Together Seed Fund (A.C.H.); MRC King’s College London Doctoral Training Partnership in Biomedical Sciences grant MR/N013700/1 (C.G., S.G.); National Institute for Health and Care Research (NIHR) Academic Clinical Lectureship (A.D. and Y.W.); NIHR Biomedical Research Centre for the Infectious Diseases Biobank (J.C. and C.M.); Rosettes Trust grant 106292/2/14/Z (A.C.H., P.V., I.Z., and V.S.); Wellcome Trust Clinical Research Career Development Fellowship Award 220589/22057 (Y.W.); and UK COVID-ImmunoTherapy Consortium grant C34399/A20265 (A.C.H.).

Author affiliations: *Peter Gorer Department of Immunobiology, School of Immunology and Microbial Sciences, King’s College London, London, SE1 9RT, United Kingdom; **Immunosurveillance Laboratory, The Francis Crick Institute, London, NW1 1AT, United Kingdom; †Breast Cancer Now Research Unit, King’s College London, London, SE1 9RT, United Kingdom; ‡Department of Medical Oncology, Guy’s and St. Thomas’ NHS Foundation Trust, London, SE1 9RT, United Kingdom; §UCL Cancer Institute, University College London, London, WClE 6DD, United Kingdom; ¶Roche Diagnostics Solutions, Pleasanton, CA, 94588; ∥Department of Plastic and Reconstructive Surgery, Royal Free NHS Foundation Trust, London, NW3 2QG, United Kingdom; #School of Hygiene & Tropical Medicine, London, WC1E 7HT, United Kingdom; #Department of Infectious Diseases, School of Immunology & Microbial Sciences, King’s College London, London, SE1 9RT, United Kingdom; *Regeneration Group, Wolfson Centre for Age-Related Diseases, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London, SE5 8AF, United Kingdom; ††Research and Development Department, Guy’s and St. Thomas’ NHS Foundation Trust, London, SE1 7EH, United Kingdom; ‡‡Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy’s and St Thomas’ NHS Foundation Trust, London, SE1 7EH, United Kingdom; and §§Infectious Diseases Biobank, Department of Infectious Diseases, School of Immunology and Microbial Sciences, King’s College London, London, SE1 9RT, United Kingdom.
