

## Heterogeneous immunological recovery trajectories revealed in post-acute COVID-19

— [Source link](#) 

Yu-Ru Su, Yu-Ru Su, Dan Yuan, Dan Yuan ...+59 more authors

**Institutions:** Fred Hutchinson Cancer Research Center, Institute for Systems Biology, University of Washington, Stanford University ...+4 more institutions

**Published on:** 20 Mar 2021 - medRxiv (Cold Spring Harbor Laboratory Press)

**Topics:** Convalescence and Immune system

Related papers:

- [Predicting the severity of disease progression in COVID-19 at the individual and population level: A mathematical model](#)
- [Common and rare variant analyses combined with single-cell multiomics reveal cell-type-specific molecular mechanisms of COVID-19 severity](#)
- [Mapping the human genetic architecture of COVID-19 by worldwide meta-analysis](#)
- [Molecular Phenomic Approaches to Deconvolving the Systemic Effects of SARS-CoV-2 Infection and Post-acute COVID-19 Syndrome](#)
- [Mapping the human genetic architecture of COVID-19.](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/heterogeneous-immunological-recovery-trajectories-revealed-4igc7cyubv>

# Heterogeneous immunological recovery trajectories revealed in post-acute COVID-19

Yapeng Su<sup>1,2,3,\*†</sup>, Dan Yuan<sup>1,4†</sup>, Daniel G. Chen<sup>1†</sup>, Kai Wang<sup>1</sup>, Jongchan Choi<sup>1</sup>, Chengzhen L. Dai<sup>1</sup>, Sunga Hong<sup>1</sup>, Rongyu Zhang<sup>1,4</sup>, Jingyi Xie<sup>1</sup>, Sarah Li<sup>1</sup>, Kelsey Scherler<sup>1</sup>, Ana Jimena Pavlovitch-Bedzyk<sup>5</sup>, Shen Dong<sup>6</sup>, Christopher Lausted<sup>1</sup>, Rachel H. Ng<sup>1,4</sup>, Inyoul Lee<sup>1</sup>, Shannon Fallen<sup>1</sup>, Sergey A. Kornilov<sup>1</sup>, Priyanka Baloni<sup>1</sup>, Venkata R. Duvvuri<sup>1</sup>, Kristin G. Anderson<sup>3,7</sup>, Jing Li<sup>5</sup>, Fan Yang<sup>8</sup>, Clifford Rostomily<sup>1</sup>, Pamela Troisch<sup>1</sup>, Brett Smith<sup>1</sup>, Jing Zhou<sup>9</sup>, Sean Mackay<sup>9</sup>, Kim Murray<sup>1</sup>, Rick Edmark<sup>1</sup>, Lesley Jones<sup>1</sup>, Yong Zhou<sup>1</sup>, Lee Rowen<sup>1</sup>, Rachel Liu<sup>1</sup>, William Chour<sup>1</sup>, William R. Berrington<sup>10,11</sup>, Julie A Wallick<sup>10,11</sup>, Heather A Algren<sup>10,11</sup>, the ISB-Swedish COVID19 Biobanking Unit, Terri Wrin<sup>12</sup>, Christos J. Petropoulos<sup>12</sup>, Wei Wei<sup>1</sup>, Nathan D. Price<sup>1</sup>, Naeha Subramanian<sup>1,13</sup>, Jennifer Hadlock<sup>1</sup>, Andrew T. Magis<sup>1</sup>, Antoni Ribas<sup>14</sup>, Lewis L. Lanier<sup>15</sup>, Scott D. Boyd<sup>8</sup>, Jeffrey A. Bluestone<sup>6</sup>, Leroy Hood<sup>1,11</sup>, Raphael Gottardo<sup>2,16,17</sup>, Philip D. Greenberg<sup>3,7</sup>, Mark M. Davis<sup>5,18,19</sup>, Jason D. Goldman<sup>10,11,20\*</sup>, James R. Heath<sup>1,4\*</sup>

1Institute for Systems Biology, Seattle, WA, USA;

2Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

3Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

4Department of Bioengineering, University of Washington, Seattle, WA, USA;

5Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, CA, USA;

6Diabetes Center, University of California, San Francisco, CA, USA;

7Departments of Immunology and Medicine, University of Washington, Seattle, WA, USA;

8Department of Pathology, Stanford University, Stanford, CA, USA;

9Isoplexis Corporation, Branford, CT, USA;

10Swedish Center for Research and Innovation, Swedish Medical Center, Seattle, WA, USA;

11Providence St. Joseph Health, Renton, WA, USA;

12Monogram Biosciences, South San Francisco, CA, USA;

13Department of Global Health, and Department of Immunology, University of Washington, Seattle, WA, USA;

14Department of Medicine, University of California, Los Angeles, and Parker Institute for Cancer Immunotherapy, Los Angeles, California, USA;

15Department of Microbiology and Immunology, University of California, San Francisco, and Parker Institute for Cancer Immunotherapy, San Francisco, CA, USA;

16Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA;

17Department of Statistics, University of Washington, Seattle, WA, USA;

18Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA;

19The Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA;

20Division of Allergy & Infectious Diseases, University of Washington, Seattle, WA, USA.

† These authors contributed equally to this work. \* Corresponding authors.

## **Abstract:**

The immunological picture of how different patients recover from COVID-19, and how those recovery trajectories are influenced by infection severity, remain unclear. We investigated 140 COVID-19 patients from diagnosis to convalescence using clinical data, viral load assessments, and multi-omic analyses of blood plasma and circulating immune cells. Immune-phenotype dynamics resolved four recovery trajectories. One trajectory signals a return to pre-infection healthy baseline, while the other three are characterized by differing fractions of persistent cytotoxic and proliferative T cells, distinct B cell maturation processes, and memory-like innate immunity. We resolve a small panel of plasma proteins that, when measured at diagnosis, can predict patient survival and recovery-trajectory commitment. Our study offers novel insights into post-acute immunological outcomes of COVID-19 that likely influence long-term adverse sequelae.

## Introduction

While the COVID-19 pandemic has resulted in over 2.3 million deaths worldwide, approximately 98% of infected individuals survive the infection. However, many survived patients either do not fully recover or only slowly regain their pre-infection baseline health status (1). Post-acute COVID-19 defines patients with symptoms extending beyond three weeks from initial onset, and chronic as beyond 12 weeks (2, 3). Although acute disease has been deeply characterized in several studies (4–10), most studies on convalescent patients have been centered on development and duration of immunological memory (11–14), less attention has been paid to capturing more comprehensive immunological views of the diverse recovery trajectories experienced by COVID-19 patients. However, such thorough longitudinal immunophenotyping may provide key insights for understanding long-term effects of COVID-19.

We investigated the recovery of 140 COVID-19 patients representing the full range of infection severities, through deep, multi-omic analyses of serial blood draws compared against healthy controls (Fig. 1A, table S1). Blood samples were collected at clinical diagnosis (T1, diagnosis), around 1 week later during acute disease (T2, acute), and 2-3 months later (T3, convalescent) when almost all patients had recovered from symptoms of acute disease. Each blood draw is analyzed for signatures of protective antibody immunity, large panels of plasma proteins and metabolites, and single-cell multi-omic analyses of peripheral blood mononuclear cells (PBMCs). Each blood draw was also paired with nasal swab measures of viral load and plasma viral RNAemia. These multi-omic and functional datasets were integrated within the context of electronic health records (EHR) of the same patients to connect clinical history with immunological recovery outcomes (Fig. 1A). We were able to stratify each studied patient into one of four COVID-19 recovery trajectories. Only one trajectory describes a return to a pre-infection, healthy baseline by T3.

## Results

### Peripheral immune cell composition dynamics reveal four distinct recovery trajectories

We used single-cell transcriptomes of PBMCs collected from all patients (all timepoints) to identify major immune cell classes; visualized as a two-dimensional projection via uniform manifold approximation and projection (UMAP) (15) (Fig. S1A). Each class was further classified into subtypes based on global transcriptomic profiles (Fig. S1A). Immune cell compositions widely varied across patients within and between time points (table S2). We utilized these composition dynamics to classify patients into four immunological trajectory groups (Fig. 1B). Notably, while older patients are at higher risk of fatal COVID-19 (both in general (16) and in this cohort), age differences were only significant between groups 2 and 3 (Fig. S1B). The groupings independently reflected COVID-19 infection severity. Groups 2 and 4 were enriched for more severe COVID-19 patients, as defined by the WHO Ordinal Scale Score (WOS) (17), relative to groups 1 and 3 (Fig. S1B). This underscores disease severity as a major contributor to a patient's immunological outcome at convalescence, although other factors exist.

Recursive partitioning of data from patient EHR (co-morbidities, medications, infection severity, etc.) was analyzed for clinical determinants of the immunological groupings (Fig. 1C, Supplemental Methods). These clinical metrics did not resolve differences between patient groups 2 and 4. The main differentiator was low levels of minimum blood oxygen (<95) were observed in groups 2 and 4. Age was the one variable that further resolved patient groupings. This analysis suggests clinical metrics track with immunological groupings.

Additionally, post-COVID-19 healthcare encounters for groups 2 and 4 were notable for respiratory symptoms and signs including fatigue, dyspnea, dyspnea on exertion, nausea/vomiting, the need for supplemental oxygen and/or ongoing mechanical ventilation (47% in Group 2 and 40% in Group 4). However, observations were only available for patients who had additional healthcare visits between T2 and T3. Additional research is needed to characterize patterns of longitudinal sequelae, compare results with SARS-CoV-2 negative reference populations, and investigate underlying etiology.

Viral load from qPCR analysis of nasal swabs correlates with COVID-19 infection severity (18) and mortality (19). RNAemia (detection of viral-RNA in plasma) also correlates with infection severity (20), but is only detected in a subset of patients. RNAemia assays discriminated between fatal (D), severe (WOS=4-7) and mild (WOS=1-3) disease at T1 and T2 (Fig. 1D).

### **Plasma multi-omics integrated with clinical data reveal divergent outcomes across patient groups**

Figure. 1E displays kinetic changes, per patient group, in five proteins reported to increase with acute infection severity (4, 9). Here, the y-axis represents natural log-fold changes above or below the value (set to 0) expected for matched non-COVID-19 healthy controls (Supplemental Methods). Notably, all groups show decreasing trends implying COVID-19 recovery (Fig. 1E top and middle). In contrast, many other measured proteins behave otherwise (table S5). We grouped proteins into specific gene ontology (GO) processes shared across all groups (Supplemental Methods) collectively representing  $\geq 200$  proteins. “Developmental Processes”, for example, reflects construction of an immune response during acute disease, and post-infection tissue repair. For the two mildly infected groups, group 3 exhibited an infection response at T1 that returned to near-healthy by T3, while group 1 exhibited the opposite trend (Fig. 1E bottom). Groups 2 and 4 exhibited strong responses at T1 that remained high at T3 (Fig. 1E bottom). This comprehensive proteomic view suggests continued engagement of immune processes in groups 1, 2 and 4.

Significant metabolic changes are also observed across patient groups at T3, including elevated pro-inflammatory metabolites (21, 22) in group 4, and immunosuppressive metabolites in group 2 (23–26) (Fig. S1C, table S6).

We explored for relationships between plasma analytes (proteins and metabolites) with clinical measures at T2 (Supplemental Methods) to identify surrogate markers for clinical metrics, which are typically less available at T3. Unsupervised clustering of correlation patterns grouped clinical measures via function (e.g. clotting metrics-row-cluster 1) (row clusters of Fig. 1F, Fig. S1D) and resolved correlated clusters of plasma analytes (columns of Fig. 1F, Fig. S1D, table S7). For example, plasma cluster 4 negatively correlates with clinical-severity-metrics, while clusters 8 and 9, which include IL-6 and TNF (table S7), positive-correlated with severity. Clusters 5 and 7 correlated with clinical metrics of liver (ALT, AST) and kidney (creatinine, BUN) damage, respectively (Fig. 1F), and were elevated in patient groups 1, 2 and 4 at T3 (Fig. 1G). These elevations suggest possible continued kidney and liver damage in some patients at T3, and again imply a slow recovery for certain patients.

### **Patients with history of severe COVID-19 can have cytotoxic and proliferative T cells in convalescence**

We analyzed for CD8+ and CD4+ T cell phenotype differences across the patient groups. We projected CD8+ and CD4+ T cell sc-CITE-seq data onto separate UMAP representations to identify distinct T cell phenotypes (Fig. 2A, Fig. S2A-C, S3A-C). Clonal expansion in group 2 was high for CD8+ and CD4+ T cells

at T2 and T3 (Fig. S2D, S3D). Two T cell phenotypes previously reported to positively-correlate with disease severity(4, 8) remained unexpectedly enriched in some convalescent patients (Fig. 2B, Fig. S2A-C, E, S3A-C, E, F). The first are cytotoxic CD8+ and CD4+ T cells, both with upregulated CD45RA protein and high clonal expansion levels. The second is a hybrid (proliferative-exhausted) phenotype with elevated levels of proliferative (MKI67) and exhaustion (e.g. LAG3) transcripts (4, 8). Notably, groups 2 and 4 exhibited elevated cytotoxic T cell percentages even at T3. Cytotoxic CD4+ T cells should rapidly contract post-disappearance of antigens (27), and so their persistence at T3 suggests an ongoing T cell response (or an anomalously slow contraction) 2-3 months after infection.

Patient group 4 exhibited elevated fractions of the proliferative-exhausted phenotypes at T3 (Fig. 2B). Interestingly, group 4 also showed reduced cytotoxic CD4+ T cell fractions during acute infection (T2) relative to group 2 (Fig. 2B). Continued T cell proliferation and compromised early (T2) T cell responses suggests potential delayed antigen clearance. Group 4 also exhibited repressed levels of T follicular helper (TFH) cells at T3 (Fig. 2B). T-FH cells facilitate somatic hypermutation and isotype switching in B cells to promote humoral immunity. We return to this issue in our B cell analysis below. Group 2 is distinguished by upregulated Th17 and Treg cells at T3 (Fig. S3G). Upregulated Th17 cells can be linked to autoimmunity risks (28) and upregulated Treg counts, especially activated Treg cells (Fig. S3G), may be controlling the persistent elevation of cytotoxic T cells in group 2 at T3.

### **Convalescent patients exhibit higher T cell polyfunctionality than unexposed healthy controls**

To evaluate the responsiveness of T cells in recovered patients, we used single-cell secretome technology (29–31) to measure 32 secreted proteins from CD8+ and CD4+ T cells following stimulation. A polyfunctional strength index (PSI) was calculated, reflecting the product of the numbers of different proteins secreted and the secretion copy numbers. Relative to healthy donor T cells, the PSI at T3 was elevated in groups 1, 2 and 4 for both CD8+ and CD4+ T cells (Fig. 2C). This high PSI suggests that T cells at convalescence are more primed to secrete multiple cytokines during a possible recall response. Many of those cytokines possess effector (cytotoxic) functions (Fig. S2H, S3H), but other functional signatures were also detected. As examples, CD4+ T cells from groups 2 and 4 secreted relatively high levels of Th1-like cytokines (e.g. IFN- $\gamma$ ), while CD4+ T cells from group 2 patients uniquely secreted elevated levels of Th2-(IL-4), and Th17-(IL-17A) like cytokines (Fig. S3H). Simultaneous elevation of Th1, Th2 and Th17 within group 2 is consistent with the sc-CITE-seq and suggests a more diverse CD4+ T cell response upon stimulation.

### **Clonal and cell-state-transition dynamics reveal divergent groups of TCRs and an early-transcriptome signature associated with memory formation versus clonal contraction**

TCR gene sequences provide molecular barcodes to track clonal frequency and transcriptomic alteration from early infection to convalescence. We identified that clonally dominant TCRs at T3 are different from those dominant at T2 (Fig. 2D). We investigated this by separately clustering CD8+ and CD4+ T cell TCRs via their clonal frequencies across distinct transcriptomic phenotypes and time (Fig. 2D, Fig. S4A, table S8, Supplemental Methods). For CD8+ T cells, TCR clusters 3 and 4 were enriched within the effector-memory pool but cluster 3 TCRs were dominant during early infection (T2) and contracted in convalescence (T3), whereas cluster 4 TCRs were clonally dominant at T3 but not T2, for all patient groups (Fig. S4B). Both clusters 3 and 4 contain likely SARS-CoV-2-specific TCRs as revealed by GLIPH2 analysis with published SARS-CoV-2-specific TCRs (32, 33) (Fig. 2F). Certain SARS-CoV-2 epitope

specificities are shared across both TCR clusters, while others are cluster specific (Fig. S4C, table S9). Similarly divergent TCR clusters were also revealed in CD4+ T cells (Fig. 2D, F, Fig. S4C).

Using TCRs as lineage-tracing barcodes, we mined for early transcriptomic signatures to predict these TCR dynamics and subsequent cell fates. Specifically, within cytotoxic CD8+ T cells in T2, we queried for transcriptomic differences associated with clonotypes undergoing conversion to memory T cells at T3 versus clonal contraction at T3 (Fig. 2E). “Memory-precursor” cytotoxic T cells at T2 showed biased activation of memory-like genes (e.g. IL7R(34), GZMK(35)), as well as genes that inhibit inflammation or prevent T cell over-activation (e.g. DUSP239, JUNB40) (Fig. 2E bottom-left blue-bars). By contrast, T2 cytotoxic T cells destined for clonal contraction had upregulated genes associated with effector functions (e.g. GZMB, PRF1) and inflammatory-responses (Fig. 2E bottom-left orange-bars). Thus, cytotoxic CD8+ T cells in early-stage COVID-19 with hyper-activation signatures may eventually contract, while subpopulations with early memory-biased signatures will become dominant in the memory pool later. However even at T3, the cytotoxic pool was replenished with new clones (Fig. 2D cluster 2). Similarly, for CD4+ T cells, cytotoxic CD4+ T cells at T2 that upregulate inflammatory-response pathways (e.g. LPS stimulatory response) tended to clonally contract (Fig. 2E bottom-right orange-bars). CD4+ clonotypes that had more regulatory signal responses (e.g. IL-10 signaling) tended towards clonal expansion at T3 (Fig. 2E bottom-right blue-bars). Such early-transcriptomic signatures may potentially provide markers for identifying memory forming TCRs versus short-lived early-protective TCRs. In addition, this analysis implies that some cytotoxic, SARS-CoV-2 specific CD8+ and CD4+ T cell clonotypes continue to emerge at T3, suggesting that viral-antigen presentation may persist 2-3 months post-infection in some patients.

### **Anti-SARS-CoV-2 antibody titers observed in convalescent patients correlated with early infection severity**

Anti-SARS-CoV-2 antibody responses were elevated for all convalescent patients compare to unexposed healthy donors. We measured IgG, IgA and IgM antibody titers against spike protein, receptor-binding domain (RBD) of spike, and nucleocapsid protein at T3 (Fig. 3A-C, Fig. S5G). Higher IgG and IgA antibody titers against spike or RBD were observed for groups 2 and 4 (Fig. 3A-C), consistent with severe infection histories in those groups (38, 39). A pseudo-virus neutralization assay (Fig. 3D) showed that neutralizing antibody levels decreased between T2 and T3 in all groups (Fig. 3D middle), and strongly correlated ( $r=0.77$ ,  $p=2.2e-16$ ) with anti-RBD IgG titers (Fig. 3D right), as reported (40, 41).

### **Distinct longitudinal patterns of B cell activation and affinity maturation within patients with severe infection histories**

Clustering of B cell single-cell transcriptomes resolved naïve (IgD+CD20+), intermediate (CD27<sup>low</sup>IgD<sup>int</sup>), memory (CD27+CD20+), and antibody-secreting (ASCs, CD20-CD38+) (Fig. 5SA-E, top) phenotypes. ASCs were elevated in group 4 at T3 (Fig. 3E), indicating prolonged antigen-driven B cell activation and differentiation. Interestingly, proliferative T cells are also higher in group 4 at T3 (Fig. 2B).

Virus-specific memory B cells mature affinity for antigen through accumulation of somatic hypermutations (SHM) (42). For most patients, SHM levels were stable over time, (Fig. 3F) as reported (43). However group 2 exhibited anomalously high overall SHM levels in the memory B cell pool at T3 (Fig. 3F, Fig. S5F), while also showing a preference towards an isotype-switched phenotype (IGHG1) (Fig. 3G). These observations suggest a robust germinal-center (GC)-associated immune response. By



contrast, group 4 patients exhibited a high percentage of IGHM memory B cells relative to group 2 (Fig. 3F, G), perhaps suggesting an extrafollicular pathway for B cell differentiation. This can be related to the low percentage of circulating TFH cells in group 4 at T3 (Fig. 2B) (44). However, the ultimate effect on neutralizing antibodies is unclear (Fig. 3D). These distinctions between groups 2 and 4 suggest that factors besides disease severity may influence B cell memory formation.

Of note, we identified convergent BCR sequences with no/low levels of SHM that were reported to be specific for SARS-CoV-2 spike protein (40, 45, 46) in patients at T1 and T2 (Fig. 3H), suggesting that SARS-CoV-2-specific B cells can be generated from naive B cell pool (40, 43, 45, 46).

### **Innate immune cell classes in convalescent patients exhibit elevated functionality compared to unexposed healthy donors**

We analyzed polyfunctional responsiveness of monocytes and NK cells at T3 following stimulation. We profiled monocytes (Fig. S6A-C) and identified that all recovered patients showed higher levels of polyfunctionality than healthy donors (Fig. S6D, E). For NK cells, group 2 patients showed the highest effector cytokine polyfunctional responses (Fig. S7E, F). This may be associated with elevated percentages of memory NK cells (KLRC2<sup>high</sup>GZMH<sup>high</sup>FCER1G<sup>low</sup>PLZF<sup>low</sup>) in this group at convalescence (Fig. S7A-D). Elevated functional responsiveness of innate cells in convalescent patients suggests a memory-like behavior of innate immunity, but its potential protective role warrants further investigation.

### **Cross-dataset correlations resolve an orchestrated immune response throughout acute infection to convalescence**

To provide an integrated view of the distinct immunological trajectories taken by different patient groups, we analyzed pairwise correlations across all metrics; including functional humoral immunity data, clinical measurements, sc-CITE-seq, single-cell secretome data, and plasma-omic data sets across T2 and T3. Hierarchical clustering of these measurements yielded 14 sets, each defined by similar cross-correlation patterns (Fig. 4A). For these 14 sets, similar functions (e.g. polyfunctional responses of monocytes) tended to cluster together into the same set (Fig. 4A right), suggesting coordination within specific immune cell classes. Relative levels of metrics across patient groups are summarized as a heatmap (Fig. 4A top)

The matrix reveals interesting correlation patterns. Several measures of monocyte PSI at T3 are uniquely high in group 3, and are anti-correlated with whether patients exhibited a cough at T2. CD4+ T cell functionalities (Set 13, 14) exhibited strong correlations with CD8+ T cell PSI metrics (Set 10, 11) and various antibody titers (Set 9) (Fig. 4A, Fig. S8), likely reflecting the key role of CD4+ T cells in orchestrating different arms of immunity. Thus, this correlation matrix illustrates how different immune components coordinate within specific patient groups and further connects that coordination to specific clinical observations.

Further inspection also revealed that levels of plasma protein clusters (C1, C4, C8 and C9) appear to discriminate between patient groups and have extensive correlations with T cell responsiveness, antibody titers and clinical measures (Fig. 4A). This prompted us to identify plasma protein markers to predict immunological recovery trajectory commitment.

## **Machine learning model resolves five plasma protein signature at disease diagnosis that accurately predicts patient immunological trajectory**

We utilized machine learning to explore plasma proteins that, when measured at diagnosis (T1), were predictive of patient survival from COVID-19 as well as which of the four distinct immunological trajectories the patient would follow (Fig. 4B). We resolved a logistic-regression model that utilized measured levels of five proteins at T1 for predicting all five different outcomes (Supplemental Methods). Pairwise comparisons between outcomes yielded accurate predictions with mean area-under-curve (AUC) values of 0.72-0.94 for validation sets (Fig. 4B middle, Fig. S9A). Each protein measurement is weighted differently per pairwise comparison (Fig. 4B right). The ability of this five-protein panel to resolve all five immunological outcomes suggests that patient trajectories can be predicted at initial clinical diagnosis and underscores the importance of connecting long-term immunological outcomes with early-stage infection status.

### **Discussion**

This comprehensive profiling of blood plasma and peripheral immune cells of 140 COVID-19 patients is shown as fundamental for categorizing and predicting heterogeneous immunological trajectories during post-disease recovery of COVID-19 (Fig. S10).

Unsupervised clustering of single-cell data revealed four different paths towards recovery, two of which associate with patients who experienced mild COVID-19 (groups 1 and 3), while groups 2 and 4 associate with severe infection histories. Groups 1, 2 and 4 exhibit elevated levels of proteins at T3 that correlate with clinical measures of kidney stress, liver damage and clotting problems (Fig. 1F, G, Fig. S1D). Only patient group 3, representing 36% of our cohort, followed a path of potential return to a healthy baseline at T3. Notably, these patient groups were not distinguished by pre-existing chronic conditions or medications, but could partially be defined by a combination of blood oxygenation level at early-stage disease and patient age (Fig. 1C).

Groups are further distinguished by adaptive and innate immunity. Group 4 exhibited decreased isotype switching and SHM within memory B cells and low levels of TFH cells, pointing to an extrafollicular B cell maturation process potentially resulting from disruption of germinal center formation (47). The differences between group 2 and 4 imply that severe infections don't guarantee high-affinity class-switched memory B cells, and may suggest the necessity for vaccination of a defined subset of previously infected patients.

Groups 2 and 4 also had elevated fractions of cytotoxic CD8+ and CD4+ T cells at T3 (Fig. 2B), and high T cell polyfunctionality with secreted proteins skewed towards effector molecules (Fig. 2C, Fig. S2G, H, S3H). For both groups, many cytotoxic T cells at T3 were found to exhibit SARS-CoV-2-specific TCR sequences (Fig. 2F). Group 4 patients also exhibited high levels of proliferating T cells. The persistence of cytotoxic and proliferating T cells at T3 may suggest continued exposure to viral antigens. Persistent cytotoxic CD4+ T cells may point to potential immune-pathological risks since they can kill antigen presenting cells (48).

Elevated responsiveness of innate immune cells in recovered patients, together with the elevation of memory-like NK cells in group 2 at T3, implies "memory-like innate immunity" may exist in some patients, but its role of protections from future infections is worth further investigation.

How antiviral T cell populations give rise to T cell subsets with short- and long-lived cell fates is still a fundamental puzzle of T cell immunity (49, 50). Surprisingly, we find that the expanded T cell clones at T2 are different from those at T3. Using TCRs to lineage-trace clonotypes, we found that clones that contract by T3 had previously exhibited elevated markers of over-activation, while T3-expanded clones express genes that inhibit inflammation or prevent T cell over-activation. Whether these clonal dynamics represent more general T cell responses to infection in humans is an interesting question for future study. Further, the fact that new cytotoxic clonotypes continue to emerge at T3 suggests potential ongoing immune responses and may imply incomplete viral clearance in some patients.

Previous studies have mainly focused on biomarkers predictive of patient survival (51). We find that patient survival, as well as which of the four recovery trajectories a patient might take, can all be resolved as early as initial diagnosis, using only a panel of five plasma proteins. Such a biomarker panel may be useful in clinical settings, although further validation on a larger cohort is needed.

Our analyses provide a comprehensive examination of the diverse immunological recovery trajectories for COVID-19. The connections between post-acute outcomes with early infection raise the possibility that interventions early-on may be beneficial in reducing certain outcomes after acute disease. Our analyses show persistence of immune engagement with SARS-CoV-2 and may provide a framework to understand the phenomenon of “long COVID”. Further research is needed to correlate the patient-described and clinician-confirmed symptoms of “long COVID” with our reported immunological disease trajectories. These data may provide a knowledge base for guiding interventional trials to treat and prevent post-COVID-19 symptoms.

## References

1. C. Huang, L. Huang, Y. Wang, X. Li, L. Ren, X. Gu, L. Kang, L. Guo, M. Liu, X. Zhou, J. Luo, Z. Huang, S. Tu, Y. Zhao, L. Chen, D. Xu, Y. Li, C. Li, L. Peng, Y. Li, W. Xie, D. Cui, L. Shang, G. Fan, J. Xu, G. Wang, Y. Wang, J. Zhong, C. Wang, J. Wang, D. Zhang, B. Cao, 6-month consequences of COVID-19 in patients discharged from hospital: a cohort study. *Lancet*. 397, 220–232 (2021).
2. C. del Rio, L. F. Collins, P. Malani, Long-term health consequences of COVID-19. *JAMA*. 324, 1723–1724 (2020).
3. T. Greenhalgh, M. Knight, C. A’Court, M. Buxton, L. Husain, Management of post-acute covid-19 in primary care. *BMJ*. 370, m3026 (2020).
4. Y. Su, D. Chen, D. Yuan, C. Lausted, J. Choi, C. L. Dai, V. Voillet, V. R. Duvvuri, K. Scherler, P. Troisch, P. Baloni, G. Qin, B. Smith, S. A. Kornilov, C. Rostomily, A. Xu, J. Li, S. Dong, A. Rothchild, J. Zhou, K. Murray, R. Edmark, S. Hong, J. E. Heath, J. Earls, R. Zhang, J. Xie, S. Li, R. Roper, L. Jones, Y. Zhou, L. Rowen, R. Liu, S. Mackay, D. S. O’Mahony, C. R. Dale, J. A. Wallick, H. A. Algren, M. A. Zager, W. Wei, N. D. Price, S. Huang, N. Subramanian, K. Wang, A. T. Magis, J. J. Hadlock, L. Hood, A. Aderem, J. A. Bluestone, L. L. Lanier, P. D. Greenberg, R. Gottardo, M. M. Davis, J. D. Goldman, J. R. Heath, Multi-omics resolves a sharp disease-state shift between mild and moderate COVID-19. *Cell*. 183, 1479-1495.e20 (2020).
5. J. Schulte-Schrepping, N. Reusch, D. Paclik, K. Baßler, S. Schlickeiser, B. Zhang, B. Krämer, T. Krammer, S. Brumhard, L. Bonaguro, E. De Domenico, D. Wendisch, M. Grasshoff, T. S. Kapellos, M. Beckstette, T. Pecht, A. Saglam, O. Dietrich, H. E. Mei, A. R. Schulz, C. Conrad, D. Kunkel, E. Vafadarnejad, C.-J. Xu, A. Horne, M. Herbert, A. Drews, C. Thibeault, M. Pfeiffer, S. Hippenstiel, A. Hocke, H. Müller-

Redetzky, K.-M. Heim, F. Machleidt, A. Uhrig, L. Bosquillon de Jarcy, L. Jürgens, M. Stegemann, C. R. Glösenkamp, H.-D. Volk, C. Goffinet, M. Landthaler, E. Wyler, P. Georg, M. Schneider, C. Dang-Heine, N. Neuwinger, K. Kappert, R. Tauber, V. Corman, J. Raabe, K. M. Kaiser, M. T. Vinh, G. Rieke, C. Meisel, T. Ulas, M. Becker, R. Geffers, M. Witzenrath, C. Drosten, N. Suttorp, C. von Kalle, F. Kurth, K. Händler, J. L. Schultze, A. C. Aschenbrenner, Y. Li, J. Nattermann, B. Sawitzki, A.-E. Saliba, L. E. Sander, A. Angelov, R. Bals, A. Bartholomäus, A. Becker, D. Bezdan, E. Bonifacio, P. Bork, T. Clavel, M. Colome-Tatche, A. Diefenbach, A. Dilthey, N. Fischer, K. Förstner, J.-S. Frick, J. Gagneur, A. Goesmann, T. Hain, M. Hummel, S. Janssen, J. Kalinowski, R. Kallies, B. Kehr, A. Keller, S. Kim-Hellmuth, C. Klein, O. Kohlbacher, J. O. Korbel, I. Kurth, M. Landthaler, Y. Li, K. Ludwig, O. Makarewicz, M. Marz, A. McHardy, C. Mertes, M. Nöthen, P. Nürnberg, U. Ohler, S. Ossowski, J. Overmann, S. Peter, K. Pfeffer, A. R. Poetsch, A. Pühler, N. Rajewsky, M. Ralser, O. Rieß, S. Ripke, U. Nunes da Rocha, P. Rosenstiel, A.-E. Saliba, L. E. Sander, B. Sawitzki, P. Schiffer, E.-C. Schulte, J. L. Schultze, A. Sczyrba, O. Stegle, J. Stoye, F. Theis, J. Vehreschild, J. Vogel, M. von Kleist, A. Walker, J. Walter, D. Wiczorek, J. Ziebuhr, Severe COVID-19 is marked by a dysregulated myeloid cell compartment. *Cell* (2020), doi:<https://doi.org/10.1016/j.cell.2020.08.001>.

6. A. Silvin, N. Chapuis, G. Dunsmore, A.-G. Goubet, A. Dubuisson, L. Derosa, C. Almiere, C. Hénon, O. Kosmider, N. Droin, P. Rameau, C. Catelain, A. Alfaro, C. Dussiau, C. Friedrich, E. Sourdeau, N. Marin, T.-A. Szwebel, D. Cantin, L. Mouthon, D. Borderie, M. Deloger, D. Bredel, S. Mouraud, D. Drubay, M. Andrieu, A.-S. Lhonneur, V. Saada, A. Stoclin, C. Willekens, F. Pommeret, F. Griscelli, L. G. Ng, Z. Zhang, P. Bost, I. Amit, F. Barlesi, A. Marabelle, F. Pène, B. Gachot, F. André, L. Zitvogel, F. Ginhoux, M. Fontenay, E. Solary, Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. *Cell* (2020), doi:<https://doi.org/10.1016/j.cell.2020.08.002>.

7. P. S. Arunachalam, F. Wimmers, C. K. P. Mok, R. A. P. M. Perera, M. Scott, T. Hagan, N. Sigal, Y. Feng, L. Bristow, O. Tak-Yin Tsang, D. Wagh, J. Collier, K. L. Pellegrini, D. Kazmin, G. Alaaeddine, W. S. Leung, J. M. C. Chan, T. S. H. Chik, C. Y. C. Choi, C. Huerta, M. Paine McCullough, H. Lv, E. Anderson, S. Edupuganti, A. A. Upadhyay, S. E. Bosinger, H. T. Maecker, P. Khatri, N. Roupheal, M. Peiris, B. Pulendran, Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science* (80-. ), eabc6261 (2020).

8. D. Mathew, J. R. Giles, A. E. Baxter, D. A. Oldridge, A. R. Greenplate, J. E. Wu, C. Alanio, L. Kuri-Cervantes, M. B. Pampena, K. D'Andrea, S. Manne, Z. Chen, Y. J. Huang, J. P. Reilly, A. R. Weisman, C. A. G. Ittner, O. Kuthuru, J. Dougherty, K. Nzingha, N. Han, J. Kim, A. Pattekar, E. C. Goodwin, E. M. Anderson, M. E. Weirick, S. Gouma, C. P. Arevalo, M. J. Bolton, F. Chen, S. F. Lacey, H. Ramage, S. Cherry, S. E. Hensley, S. A. Apostolidis, A. C. Huang, L. A. Vella, M. R. Betts, N. J. Meyer, E. J. Wherry, Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science* (80-. ), eabc8511 (2020).

9. C. Lucas, P. Wong, J. Klein, T. B. R. Castro, J. Silva, M. Sundaram, M. K. Ellingson, T. Mao, J. E. Oh, B. Israelow, T. Takahashi, M. Tokuyama, P. Lu, A. Venkataraman, A. Park, S. Mohanty, H. Wang, A. L. Wyllie, C. B. F. Vogels, R. Earnest, S. Lapidus, I. M. Ott, A. J. Moore, M. C. Muenker, J. B. Fournier, M. Campbell, C. D. Odio, A. Casanovas-Massana, A. Obaid, A. Lu-Culligan, A. Nelson, A. Brito, A. Nunez, A. Martin, A. Watkins, B. Geng, C. Kalinich, C. Harden, C. Todeasa, C. Jensen, D. Kim, D. McDonald, D. Shepard, E. Courchaine, E. B. White, E. Song, E. Silva, E. Kudo, G. Deluiliis, H. Rahming, H.-J. Park, I. Matos, J. Nouws, J. Valdez, J. Fauver, J. Lim, K.-A. Rose, K. Anastasio, K. Brower, L. Glick, L. Sharma, L. Sewanan, L. Knaggs, M. Minasyan, M. Batsu, M. Petrone, M. Kuang, M. Nakahata, M. Campbell, M. Linehan, M. H.

Askenase, M. Simonov, M. Smolgovsky, N. Sonnert, N. Naushad, P. Vijayakumar, R. Martinello, R. Datta, R. Handoko, S. Bermejo, S. Prophet, S. Bickerton, S. Velazquez, T. Alpert, T. Rice, W. Khoury-Hanold, X. Peng, Y. Yang, Y. Cao, Y. Strong, R. Herbst, A. C. Shaw, R. Medzhitov, W. L. Schulz, N. D. Grubaugh, C. Dela Cruz, S. Farhadian, A. I. Ko, S. B. Omer, A. Iwasaki, Y. I. Team, Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature*. 584, 463–469 (2020).

10. X. Ren, W. Wen, X. Fan, W. Hou, B. Su, P. Cai, J. Li, Y. Liu, F. Tang, F. Zhang, Y. Yang, J. He, W. Ma, J. He, P. Wang, Q. Cao, F. Chen, Y. Chen, X. Cheng, G. Deng, X. Deng, W. Ding, Y. Feng, R. Gan, C. Guo, W. Guo, S. He, C. Jiang, J. Liang, Y. Li, J. Lin, Y. Ling, H. Liu, J. Liu, N. Liu, S.-Q. Liu, M. Luo, Q. Ma, Q. Song, W. Sun, G. Wang, F. Wang, Y. Wang, X. Wen, Q. Wu, G. Xu, X. Xie, X. Xiong, X. Xing, H. Xu, C. Yin, D. Yu, K. Yu, J. Yuan, B. Zhang, P. Zhang, T. Zhang, J. Zhao, P. Zhao, J. Zhou, W. Zhou, S. Zhong, X. Zhong, S. Zhang, L. Zhu, P. Zhu, B. Zou, J. Zou, Z. Zuo, F. Bai, X. Huang, P. Zhou, Q. Jiang, Z. Huang, J.-X. Bei, L. Wei, X.-W. Bian, X. Liu, T. Cheng, X. Li, P. Zhao, F.-S. Wang, H. Wang, B. Su, Z. Zhang, K. Qu, X. Wang, J. Chen, R. Jin, Z. Zhang, COVID-19 immune features revealed by a large-scale single cell transcriptome atlas. *Cell* (2021), doi:<https://doi.org/10.1016/j.cell.2021.01.053>.

11. L. Ni, F. Ye, M.-L. Cheng, Y. Feng, Y.-Q. Deng, H. Zhao, P. Wei, J. Ge, M. Gou, X. Li, L. Sun, T. Cao, P. Wang, C. Zhou, R. Zhang, P. Liang, H. Guo, X. Wang, C.-F. Qin, F. Chen, C. Dong, Detection of SARS-CoV-2-specific humoral and cellular immunity in COVID-19 convalescent individuals. *Immunity*. 52, 971-977.e3 (2020).

12. L. B. Rodda, J. Netland, L. Shehata, K. B. Pruner, P. A. Morawski, C. D. Thouvenel, K. K. Takehara, J. Eggenberger, E. A. Hemann, H. R. Waterman, M. L. Fahning, Y. Chen, M. Hale, J. Rathe, C. Stokes, S. Wrenn, B. Fiala, L. Carter, J. A. Hamerman, N. P. King, M. Gale Jr., D. J. Campbell, D. J. Rawlings, M. Pepper, Functional SARS-CoV-2-specific immune memory persists after mild COVID-19. *Cell*. 184, 169-183.e17 (2021).

13. J. M. Dan, J. Mateus, Y. Kato, K. M. Hastie, E. D. Yu, C. E. Faliti, A. Grifoni, S. I. Ramirez, S. Haupt, A. Frazier, C. Nakao, V. Rayaprolu, S. A. Rawlings, B. Peters, F. Krammer, V. Simon, E. O. Saphire, D. M. Smith, D. Weiskopf, A. Sette, S. Crotty, Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science* (80-. ), eabf4063 (2021).

14. Y. Peng, A. J. Mentzer, G. Liu, X. Yao, Z. Yin, D. Dong, W. Dejnirattisai, T. Rostron, P. Supasa, C. Liu, C. López-Camacho, J. Slon-Campos, Y. Zhao, D. I. Stuart, G. C. Paesen, J. M. Grimes, A. A. Antson, O. W. Bayfield, D. E. D. P. Hawkins, D.-S. Ker, B. Wang, L. Turtle, K. Subramaniam, P. Thomson, P. Zhang, C. Dold, J. Ratcliff, P. Simmonds, T. de Silva, P. Sopp, D. Wellington, U. Rajapaksa, Y.-L. Chen, M. Salio, G. Napolitani, W. Paes, P. Borrow, B. M. Kessler, J. W. Fry, N. F. Schwabe, M. G. Semple, J. K. Baillie, S. C. Moore, P. J. M. Openshaw, M. A. Ansari, S. Dunachie, E. Barnes, J. Frater, G. Kerr, P. Goulder, T. Lockett, R. Levin, Y. Zhang, R. Jing, L.-P. Ho, E. Barnes, D. Dong, T. Dong, S. Dunachie, J. Frater, P. Goulder, G. Kerr, P. Klenerman, G. Liu, A. McMichael, G. Napolitani, G. Ogg, Y. Peng, M. Salio, X. Yao, Z. Yin, J. Kenneth Baillie, P. Klenerman, A. J. Mentzer, S. C. Moore, P. J. M. Openshaw, M. G. Semple, D. I. Stuart, L. Turtle, R. J. Cornall, C. P. Conlon, P. Klenerman, G. R. Screaton, J. Mongkolsapaya, A. McMichael, J. C. Knight, G. Ogg, T. Dong, O. I. N. C.-19 R. T. cell Consortium, I. Investigators, 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 (2020).

15. E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, E. W. Newell, Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* 37, 38–44 (2019).
16. Z. Shahid, R. Kalayanamitra, B. McClafferty, D. Kepko, D. Ramgobin, R. Patel, C. S. Aggarwal, R. Vunnam, N. Sahu, D. Bhatt, K. Jones, R. Golamari, R. Jain, COVID-19 and older adults: what we know. *J. Am. Geriatr. Soc.* 68, 926–929 (2020).
17. World Health Organization, WHO R&D blueprint novel coronavirus COVID-19 therapeutic trial synopsis (2020), pp. 1–9.
18. K. V Argyropoulos, A. Serrano, J. Hu, M. Black, X. Feng, G. Shen, M. Call, M. J. Kim, A. Lytle, B. Belovarac, T. Vougiouklakis, L. H. Lin, U. Moran, A. Heguy, A. Troxel, M. Snuderl, I. Osman, P. Cotzia, G. Jour, Association of initial viral load in severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) patients with outcome and symptoms. *Am. J. Pathol.* 190, 1881–1887 (2020).
19. E. Pujadas, F. Chaudhry, R. McBride, F. Richter, S. Zhao, A. Wajnberg, G. Nadkarni, B. S. Glicksberg, J. Houldsworth, C. Cordon-Cardo, SARS-CoV-2 viral load predicts COVID-19 mortality. *Lancet. Respir. Med.* 8 (2020), p. e70.
20. C. A. Hogan, B. A. Stevens, M. K. Sahoo, C. Huang, N. Garamani, S. Gombar, F. Yamamoto, K. Murugesan, J. Kurzer, J. Zehnder, B. A. Pinsky, High frequency of SARS-CoV-2 RNAemia and association with severe disease. *Clin. Infect. Dis.* (2020), doi:10.1093/cid/ciaa1054.
21. J. 3rd Weiner, S. K. Parida, J. Maertzdorf, G. F. Black, D. Repsilber, A. Telaar, R. P. Mohny, C. Arndt-Sullivan, C. A. Ganoza, K. C. Faé, G. Walzl, S. H. E. Kaufmann, Biomarkers of inflammation, immunosuppression and stress with active disease are revealed by metabolomic profiling of tuberculosis patients. *PLoS One.* 7, e40221 (2012).
22. M. Milanski, G. Degasperi, A. Coope, J. Morari, R. Denis, D. E. Cintra, D. M. L. Tsukumo, G. Anhe, M. E. Amaral, H. K. Takahashi, R. Curi, H. C. Oliveira, J. B. C. Carvalheira, S. Bordin, M. J. Saad, L. A. Velloso, Saturated fatty acids produce an inflammatory response predominantly through the activation of TLR4 signaling in hypothalamus: Implications for the pathogenesis of obesity. *J. Neurosci.* 29, 359–370 (2009).
23. V. L. Wehbi, K. Taskén, Molecular mechanisms for cAMP-mediated immunoregulation in T cells - role of anchored protein kinase signaling units. *Front. Immunol.* 7, 222 (2016).
24. J. S. Ungerstedt, M. Blömbäck, T. Söderström, Nicotinamide is a potent inhibitor of proinflammatory cytokines. *Clin. Exp. Immunol.* 131, 48–52 (2003).
25. S. K. Ku, H. R. Cho, Y. S. Sung, S. J. Kang, Y. J. Lee, Effects of calcium gluconate on experimental periodontitis and alveolar bone loss in rats. *Basic Clin. Pharmacol. Toxicol.* 108, 241–250 (2011).
26. M. M. Essa, H. Hamdan, S. B. Chidambaram, B. Al-Balushi, G. J. Guillemin, D. M. Ojcius, M. W. Qoronfleh, Possible role of tryptophan and melatonin in COVID-19. *Int. J. Tryptophan Res.* 13 (2020), p. 1178646920951832.
27. L. F. Su, B. A. Kidd, A. Han, J. J. Kotzin, M. M. Davis, Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. *Immunity.* 38, 373–383 (2013).

28. J. C. Waite, D. Skokos, Th17 response and inflammatory autoimmune diseases. *Int. J. Inflam.* 2012, 819467 (2012).
29. Y. Lu, Q. Xue, M. R. Eisele, E. S. Sulistijo, K. Brower, L. Han, E. D. Amir, D. Pe\textquoterighter, K. Miller-Jensen, R. Fan, Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *Proc. Natl. Acad. Sci.* 112, E607--E615 (2015).
30. C. Ma, A. F. Cheung, T. Chodon, R. C. Koya, Z. Wu, C. Ng, E. Avramis, A. J. Cochran, O. N. Witte, D. Baltimore, B. Chmielowski, J. S. Economou, B. Comin-Anduix, A. Ribas, J. R. Heath, Multifunctional T-cell analyses to study response and progression in adoptive cell transfer immunotherapy. *Cancer Discov.* 3, 418–429 (2013).
31. J. Zhou, A. Kaiser, C. Ng, R. Karcher, T. McConnell, P. Paczkowski, C. Fernandez, M. Zhang, S. Mackay, M. Tsuji, CD8+ T-cell mediated anti-malaria protection induced by malaria vaccines; assessment of hepatic CD8+ T cells by SCBC assay. *Hum. Vaccin. Immunother.* 13, 1625–1629 (2017).
32. J. Glanville, H. Huang, A. Nau, O. Hatton, L. E. Wagar, F. Rubelt, X. Ji, A. Han, S. M. Krams, C. Pettus, N. Haas, C. S. L. Arlehamn, A. Sette, S. D. Boyd, T. J. Scriba, O. M. Martinez, M. M. Davis, Identifying specificity groups in the T cell receptor repertoire. *Nature.* 547, 94–98 (2017).
33. H. Huang, C. Wang, F. Rubelt, T. J. Scriba, M. M. Davis, Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat. Biotechnol.* 38, 1194–1202 (2020).
34. S. L. Colpitts, N. M. Dalton, P. Scott, IL-7 receptor expression provides the potential for long-term survival of both CD62Lhigh central memory T cells and Th1 effector cells during Leishmania major infection. *J. Immunol.* 182, 5702–5711 (2009).
35. N. Weng, Y. Araki, K. Subedi, The molecular basis of the memory T cell response: differential gene expression and its epigenetic regulation. *Nat. Rev. Immunol.* 12, 306–315 (2012).
36. R. Lang, F. A. M. Raffi, Dual-Specificity Phosphatases in Immunity and Infection: An Update. *Int. J. Mol. Sci.* 20, 2710 (2019).
37. S. Koizumi, D. Sasaki, T.-H. Hsieh, N. Taira, N. Arakaki, S. Yamasaki, K. Wang, S. Sarkar, H. Shirahata, M. Miyagi, H. Ishikawa, JunB regulates homeostasis and suppressive functions of effector regulatory T cells. *Nat. Commun.* 9, 5344 (2018).
38. E. H. Y. Lau, O. T. Y. Tsang, D. S. C. Hui, M. Y. W. Kwan, W. Chan, S. S. Chiu, R. L. W. Ko, K. H. Chan, S. M. S. Cheng, R. A. P. M. Perera, B. J. Cowling, L. L. M. Poon, M. Peiris, Neutralizing antibody titres in SARS-CoV-2 infections. *Nat. Commun.* 12, 63 (2021).
39. Y. Chen, A. Zuiani, S. Fischinger, J. Mullur, C. Atyeo, M. Travers, F. J. N. Lelis, K. M. Pullen, H. Martin, P. Tong, A. Gautam, S. Habibi, J. Bensko, D. Gakpo, J. Feldman, B. M. Hauser, T. M. Caradonna, Y. Cai, J. S. Burke, J. Lin, J. A. Lederer, E. C. Lam, C. L. Lavine, M. S. Seaman, B. Chen, A. G. Schmidt, A. B. Balazs, D. A. Lauffenburger, G. Alter, D. R. Wesemann, Quick COVID-19 healers sustain anti-SARS-CoV-2 antibody production. *Cell.* 183, 1496-1507.e16 (2020).
40. D. F. Robbiani, C. Gaebler, F. Muecksch, J. C. C. Lorenzi, Z. Wang, A. Cho, M. Agudelo, C. O. Barnes, A. Gazumyan, S. Finkin, T. Hägglöf, T. Y. Oliveira, C. Viant, A. Hurley, H.-H. Hoffmann, K. G.

- Millard, R. G. Kost, M. Cipolla, K. Gordon, F. Bianchini, S. T. Chen, V. Ramos, R. Patel, J. Dizon, I. Shimeliovich, P. Mendoza, H. Hartweger, L. Nogueira, M. Pack, J. Horowitz, F. Schmidt, Y. Weisblum, E. Michailidis, A. W. Ashbrook, E. Waltari, J. E. Pak, K. E. Huey-Tubman, N. Koranda, P. R. Hoffman, A. P. West, C. M. Rice, T. Hatziioannou, P. J. Bjorkman, P. D. Bieniasz, M. Caskey, M. C. Nussenzweig, Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature*. 584, 437–442 (2020).
41. K. Röltgen, A. E. Powell, O. F. Wirz, B. A. Stevens, C. A. Hogan, J. Najeeb, M. Hunter, H. Wang, M. K. Sahoo, C. Huang, F. Yamamoto, M. Manohar, J. Manalac, A. R. Otrelo-Cardoso, T. D. Pham, A. Rustagi, A. J. Rogers, N. H. Shah, C. A. Blish, J. R. Cochran, T. S. Jardetzky, J. L. Zehnder, T. T. Wang, B. Narasimhan, S. Gombar, R. Tibshirani, K. C. Nadeau, P. S. Kim, B. A. Pinsky, S. D. Boyd, Defining the features and duration of antibody responses to SARS-CoV-2 infection associated with disease severity and outcome. *Sci. Immunol.* 5, eabe0240 (2020).
42. A. Sokal, P. Chappert, A. Roeser, G. Barba-Spaeth, S. Fourati, I. Azzaoui, A. Vandenberghe, I. Fernandez, M. Bouvier-Alias, E. Crickx, A. B. Ferchiou, S. Hue, L. Languille, S. Baloul, F. Noizat-Pirenne, M. Luka, J. Megret, M. Ménager, J. M. Pawlotsky, S. Fillatreau, F. A. Rey, J. C. Weill, C. A. Reynaud, M. Mahévas, *bioRxiv*, in press, doi:10.1101/2020.11.17.385252.
43. C. Kreer, M. Zehner, T. Weber, M. S. Ercanoglu, L. Gieselmann, C. Rohde, S. Halwe, M. Korenkov, P. Schommers, K. Vanshylla, V. Di Cristanziano, H. Janicki, R. Brinker, A. Ashurov, V. Krähling, A. Kupke, H. Cohen-Dvashi, M. Koch, J. M. Eckert, S. Lederer, N. Pfeifer, T. Wolf, M. J. G. T. Vehreschild, C. Wendtner, R. Diskin, H. Gruell, S. Becker, F. Klein, Longitudinal isolation of potent near-germline SARS-CoV-2-neutralizing antibodies from COVID-19 patients. *Cell*. 182, 843-854.e12 (2020).
44. S. Crotty, T follicular helper cell differentiation, function, and roles in disease. *Immunity*. 41, 529–542 (2014).
45. S. J. Zost, P. Gilchuk, R. E. Chen, J. B. Case, J. X. Reidy, A. Trivette, R. S. Nargi, R. E. Sutton, N. Suryadevara, E. C. Chen, E. Binshtein, S. Shrihari, M. Ostrowski, H. Y. Chu, J. E. Didier, K. W. MacRenaris, T. Jones, S. Day, L. Myers, F. Eun-Hyung Lee, D. C. Nguyen, I. Sanz, D. R. Martinez, P. W. Rothlauf, L.-M. Bloyet, S. P. J. Whelan, R. S. Baric, L. B. Thackray, M. S. Diamond, R. H. Carnahan, J. E. Crowe, Rapid isolation and profiling of a diverse panel of human monoclonal antibodies targeting the SARS-CoV-2 spike protein. *Nat. Med.* 26, 1422–1427 (2020).
46. D. Li, R. J. Edwards, K. Manne, D. R. Martinez, A. Schäfer, S. M. Alam, K. Wiehe, X. Lu, R. Parks, L. L. Sutherland, T. H. Oguin, C. McDanal, L. G. Perez, K. Mansouri, S. M. C. Gobeil, K. Janowska, V. Stalls, M. Kopp, F. Cai, E. Lee, A. Foulger, G. E. Hernandez, A. Sanzone, K. Tilahun, C. Jiang, L. V Tse, K. W. Bock, M. Minai, B. M. Nagata, K. Cronin, V. Gee-Lai, M. Deyton, M. Barr, T. Von Holle, A. N. Macintyre, E. Stover, J. Feldman, B. M. Hauser, T. M. Caradonna, T. D. Scobey, M. A. Moody, D. W. Cain, C. T. DeMarco, T. N. Denny, C. W. Woods, E. W. Petzold, A. G. Schmidt, I.-T. Teng, T. Zhou, P. D. Kwong, J. R. Mascola, B. S. Graham, I. N. Moore, R. Seder, H. Andersen, M. G. Lewis, D. C. Montefiori, G. D. Sempowski, R. S. Baric, P. Acharya, B. F. Haynes, K. O. Saunders, *bioRxiv*, in press, doi:10.1101/2020.12.31.424729.
47. N. Kaneko, H.-H. Kuo, J. Boucau, J. R. Farmer, H. Allard-Chamard, V. S. Mahajan, A. Piechocka-Trocha, K. Lefteri, M. Osborn, J. Bals, Y. C. Bartsch, N. Bonheur, T. M. Caradonna, J. Chevalier, F. Chowdhury, T. J. Diefenbach, K. Einkauf, J. Fallon, J. Feldman, K. K. Finn, P. Garcia-Broncano, C. A. Hartana, B. M. Hauser, C. Jiang, P. Kaplonek, M. Karpell, E. C. Koscher, X. Lian, H. Liu, J. Liu, N. L. Ly, A. R.



Michell, Y. Rassadkina, K. Seiger, L. Sessa, S. Shin, N. Singh, W. Sun, X. Sun, H. J. Ticheli, M. T. Waring, A. L. Zhu, G. Alter, J. Z. Li, D. Lingwood, A. G. Schmidt, M. Lichterfeld, B. D. Walker, X. G. Yu, R. F. Padera Jr., S. Pillai, Loss of Bcl-6-expressing T follicular helper cells and germinal centers in COVID-19. *Cell* (2020), doi:10.1016/j.cell.2020.08.025.

48. D. Grogg, S. Hahn, P. Erb, CD4+ T cell-mediated killing of major histocompatibility complex class II-positive antigen-presenting cells (APC). III. CD4+ cytotoxic T cells induce apoptosis of APC. *Eur. J. Immunol.* 22, 267–272 (1992).

49. E. J. Wherry, R. Ahmed, Memory CD8 T-cell differentiation during viral infection. *J. Virol.* 78, 5535–5545 (2004).

50. S. M. Kaech, E. J. Wherry, Heterogeneity and cell-fate decisions in effector and memory CD8+ T cell differentiation during viral infection. *Immunity.* 27, 393–405 (2007).

51. L. Yan, H.-T. Zhang, J. Goncalves, Y. Xiao, M. Wang, Y. Guo, C. Sun, X. Tang, L. Jing, M. Zhang, X. Huang, Y. Xiao, H. Cao, Y. Chen, T. Ren, F. Wang, Y. Xiao, S. Huang, X. Tan, N. Huang, B. Jiao, C. Cheng, Y. Zhang, A. Luo, L. Mombaerts, J. Jin, Z. Cao, S. Li, H. Xu, Y. Yuan, An interpretable mortality prediction model for COVID-19 patients. *Nat. Mach. Intell.* 2, 283–288 (2020).

52. O. Manor, N. Zubair, M. P. Conomos, X. Xu, J. E. Rohwer, C. E. Krafft, J. C. Lovejoy, A. T. Magis, A multi-omic association study of trimethylamine N-oxide. *Cell Rep.* 24, 935–946 (2018).

53. Centers for Disease Control and Prevention, CDC’s diagnostic test for COVID-19 only and supplies (2020), (available at <https://www.cdc.gov/coronavirus/2019-ncov/lab/virus-requests.html>).

54. M. Maechler, P. Rousseeuw, C. Croux, V. Todorov, A. Ruckstuhl, M. Salibian-Barrera, T. Verbeke, M. Koller, E. L. T. Conceicao, M. Anna di Palma, Robustbase: basic robust statistics (2021), (available at <http://robustbase.r-forge.r-project.org/>).

55. F. Commo, B. M. Bot, nplr: n-parameter logistic regression (2016).

56. J. D. Goldman, K. Wang, K. Roltgen, S. C. A. Nielsen, J. C. Roach, S. N. Naccache, F. Yang, O. F. Wirz, K. E. Yost, J.-Y. Lee, K. Chun, T. Wrin, C. J. Petropoulos, I. Lee, S. Fallen, P. M. Manner, J. A. Wallick, H. A. Algren, K. M. Murray, Y. Su, J. Hadlock, J. Jeharajah, W. R. Berrington, G. P. Pappas, S. T. Nyatsatsang, A. L. Greninger, A. T. Satpathy, J. S. Pauk, S. D. Boyd, J. R. Heath, medRxiv, in press, doi:10.1101/2020.09.22.20192443.

57. S. L. Wolock, R. Lopez, A. M. Klein, Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst.* 8, 281–291.e9 (2019).

58. F. A. Wolf, P. Angerer, F. J. Theis, SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 19, 15 (2018).

59. L. McInnes, J. Healy, J. Melville, UMAP: uniform manifold approximation and projection for dimension reduction (2018) (available at <https://arxiv.org/abs/1802.03426>).

60. V. Traag, L. Waltman, N. J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities (2018), doi:10.1038/s41598-019-41695-z.

61. B. Van de Sande, C. Flerin, K. Davie, M. De Waegeneer, G. Hulselmans, S. Aibar, R. Seurinck, W. Saelens, R. Cannoodt, Q. Rouchon, T. Verbeiren, D. De Maeyer, J. Reumers, Y. Saeys, S. Aerts, A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat. Protoc.* 15, 2247–2276 (2020).
62. G. Sturm, T. Szabo, G. Fotakis, M. Haider, D. Rieder, Z. Trajanoski, F. Finotello, bioRxiv, in press, doi:10.1101/2020.04.10.035865.
63. T. M. Snyder, R. M. Gittelman, M. Klinger, D. H. May, E. J. Osborne, R. Taniguchi, H. J. Zahid, I. M. Kaplan, J. N. Dines, M. N. Noakes, R. Pandya, X. Chen, S. Elasady, E. Svejnoha, P. Ebert, M. W. Pesesky, P. De Almeida, H. O'Donnell, Q. DeGottardi, G. Keitany, J. Lu, A. Vong, R. Elyanow, P. Fields, J. Greissl, L. Baldo, S. Semprini, C. Cerchione, F. Nicolini, M. Mazza, O. M. Delmonte, K. Dobbs, R. Laguna-Goya, G. Carreño-Tarragona, S. Barrio, L. Imberti, A. Sottini, E. Quiros-Roldan, C. Rossi, A. Biondi, L. R. Bettini, M. D'Angio, P. Bonfanti, M. F. Tompkins, C. Alba, C. Dalgard, V. Sambri, G. Martinelli, J. D. Goldman, J. R. Heath, H. C. Su, L. D. Notarangelo, E. Paz-Artal, J. Martinez-Lopez, J. M. Carlson, H. S. Robins, medRxiv, in press, doi:10.1101/2020.07.31.20165647.
64. J. Ye, N. Ma, T. L. Madden, J. M. Ostell, IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Res.* 41, W34-40 (2013).
65. ImmunoMind Team, Immunarch: an R package for painless bioinformatics analysis of T-cell and B-cell immune repertoires (2019), doi:10.5281/zenodo.3367200.

## Materials and Methods

### Study subjects, WOS score and EHR data extraction

This study included data from 140 COVID-19 patients (63 males and 77 females) and 333 unexposed healthy controls. All enrolled patients provided written informed consent. De-identified proteomic and metabolomic data from matched healthy controls processed using the shared technical pooled control samples to enable batch-correction were previously collected from individuals enrolled in a wellness program (52) (Arivale, Seattle, WA). Healthy control samples for single-cell analyses were obtained from Bloodworks Northwest (Seattle, WA). Detailed information on age, sex, race, ethnicity, and disease history etc. of this patient cohort and healthy controls are listed in table S1. Disease severity was quantified using the WHO Ordinal Scale Score (WOS) (17). Clinical data for hospitalized patients were abstracted from electronic health records (EHR). Clinical lab data were extracted from the nearest time point to each blood draw. Procedures for the current study were approved by the Institutional Review Board (IRB) at Providence St. Joseph Health with IRB Study Number [STUDY2020000175] and the Western Institutional Review Board (WIRB) with IRB Study Number 20170658.

### Plasma and PBMC isolation

Plasma and PBMCs were isolated from patient whole blood as previously described (4). Briefly, plasma fractions were isolated from patient blood collected in EDTA-coated vacutainer tubes (BD, 366643). The PBMC fraction was isolated, counted, and aliquoted at 2.5 million cells/ml in CryoStor CS-10 freeze media (Biolife Solutions, 210102). The aliquoted EDTA-plasma and PBMCs were frozen at -80°C, transferred into liquid nitrogen and stored until use.

### Single-cell multi-omics assay

Chromium Single Cell Kits (10x Genomics, 1000165) were utilized to analyze the transcriptome, surface protein levels, TCR, and BCR sequences simultaneously from the same cell. Experiments were performed according to the manufacturer's instructions. Briefly, cryopreserved PBMCs were thawed and incubated with the 1X red blood cell lysis solution (Miltenyi Biotech, 130-094-183) to lyse any remaining red blood cells in the PBMC samples. Cells were stained with cell hashtag antibodies (BioLegend, 394661, 394663, 394665, 394667, 394669, 394671, 394673, 394675, 394677, 394679) and TotalSeq-C custom human antibodies (BioLegend, 99814). Stained cells were then loaded onto a Chromium Next GEM chip G (10X Genomics, 1000120). Cells were lysed for reverse transcription and complementary DNA (cDNA) amplification in the Chromium Controller (10X Genomics). The polyadenylated transcripts were reverse-transcribed inside each gel bead-in-emulsion afterward. Full-length cDNA along with cell barcode identifiers were PCR-amplified and sequencing libraries were prepared and normalized. The constructed library was sequenced on the Novaseq platform (Illumina).

### Viral load measurements

The miRNeasy kit (Qiagen) was used to isolate RNA from 100 µl of plasma or nasopharyngeal swab samples according to the manufacturer's instructions. The RNA was eluted from the membrane

with either 30  $\mu$ l or 50  $\mu$ l of RNase free water for plasma or nasopharyngeal swab samples respectively. To detect viral sequences, protocol from the CDC was followed (53), and primers were obtained from Integrated DNA Technologies (IDT). The qRT-PCR results were performed on a CFX-96 qPCR machine (Bio-Rad). Levels of SARS-CoV-2 RNA and human RNase P transcript were expressed as cycle threshold (Ct) value.

#### Plasma proteomics and metabolomics

Plasma concentrations of proteins and metabolites were measured as previously described (4). Batch-corrected proteomic and metabolomic data were further adjusted for age, sex and BMI, as well as their interactions, using a set of robust linear regression models estimated for each protein and metabolite separately using the external control sample of uninfected individuals that were selected using propensity score matching on a number of sociodemographic and comorbidity variables from a larger in-house sample. Models were fitted using the `lmrob` function from the R package `robustbase` with the 'KS2014' setting (54). Metabolite values were log<sub>2</sub> transformed prior to further analyses, while protein abundance values (NPX) were already log<sub>2</sub> scaled.

#### Single-cell multiplex secretome assay

Cryopreserved PBMCs were thawed and incubated in complete medium (RPMI 1640 (Gibco, 11875-093) containing 10% fetal bovine serum (FBS, Gibco, 26140-079), 1x of glutamax (Thermo, 35050061) and 100U/mL penicillin-streptomycin (Gibco, 15140-122)) overnight at 37°C, 5% CO<sub>2</sub>. After overnight recovery, CD4<sup>+</sup> and CD8<sup>+</sup> T cells were isolated using CD4<sup>+</sup> (Miltenyi Biotec, 130-045-101) and CD8<sup>+</sup> (Miltenyi Biotec, 130-045-201) microbeads sequentially. NK cells and Monocytes were isolated using CD56 MicroBeads, human (Miltenyi Biotec, 130-050-401) and the Pan Monocyte Isolation Kit (Miltenyi Biotec, 130-096-537), respectively.

The isolated CD4<sup>+</sup> and CD8<sup>+</sup> T cells were seeded at a density of 1x10<sup>5</sup> cells/well in a 96 well-plate and stimulated for six hours with plate-bound anti-CD3 antibodies (eBioscience, 16-0037-85, pre-coated at 10  $\mu$ g/ml overnight at 4°C) and 5  $\mu$ g/mL of soluble anti-CD28 antibodies (eBioscience, 16-0289-85,) in complete medium at 37°C, 5% CO<sub>2</sub>. The isolated NK cells were cultured for 12 hours in the presence of IL-2 (Biolegend, 589104, 10 ng/ml). The enriched monocytes were seeded at 1x10<sup>5</sup> cells/mL and stimulated with 10 ng/ml lipopolysaccharide (Sigma Aldrich, L2654) for 12 hours. After stimulation, the activated cells were collected, washed, and stained with membrane stain (included in the IsoPlexis kit), before being loaded onto the chip consisting of 12,000 chambers pre-coated with an array of 32 cytokine capture antibodies. The NK cells were resuspended in complete RPMI supplemented with PMA (Sigma, P8139, 5 ng/ml) and Ionomycin (Sigma, 10634, 500 ng/ml) and then loaded onto the IsoCode chip for the stimulation during the incubation. The chip was inserted into IsoLight for further incubation for 16 hours. Secreted cytokines were detected by a cocktail of detection antibodies followed by the fluorescent labeling. Fluorescent signals were analyzed by the IsoSpeak software to calculate the numbers of cytokine-secreting cells, the intensity level of cytokines, and polyfunctional strength index (PSI). Measured cytokines in each panel are listed as below.

Single-Cell Adaptive Immune cytokine panel including the following subsets of cytokines. Effector: Granzyme B, IFN- $\gamma$ , MIP-1 $\alpha$ , Perforin, TNF- $\alpha$ , TNF- $\beta$ ; Stimulatory: GM-CSF, IL-2, IL-5, IL-7, IL-8, IL-9, IL-12, IL-15, IL-21; Chemoattractive: CCL11, IP-10, MIP-1 $\beta$ , RANTES; Regulatory: IL-4, IL-10, IL-13, IL-22, TGF $\beta$ 1, sCD137, sCD40L; Inflammatory: IL-1 $\beta$ , IL-6, IL-17A, IL-17F, MCP-1, MCP-4.

Single-Cell Innate Immune cytokine panel including the following subsets of cytokines. Effector: IFN- $\gamma$ , MIP-1 $\alpha$ , TNF- $\alpha$ , TNF- $\beta$ ; Stimulatory: GM-CSF, IL-8, IL-9, IL-15, IL-18, TGF- $\alpha$ , IL-5; Chemoattractive: CCL11, IP-10, MIP-1 $\beta$ , RANTES, BCA-1; Regulatory: IL-10, IL-13, IL-22, sCD40L; Inflammatory: IL-1 $\beta$ , IL-6, IL-12-p40, IL-12, IL-17A, IL-17F, MCP-1, MCP-4, MIF; Growth Factors: EGF, PDGF-BB, VEGF.

## ELISAs

Briefly, 384-well plates (ThermoFisher, 464718) were coated with 10  $\mu$ L of 5  $\mu$ g/mL SARS-CoV-2 RBD (Invitrogen, RP-87678), spike (S) (Invitrogen, RP-87680), or nucleocapsid (N) (Invitrogen, RP-87707) protein in carbonate pH9.6 buffer overnight at 4°C. Plates were washed four times with wash buffer (phosphate buffered saline (PBS) containing 0.05% Tween-20) and blocked with blocking buffer (wash buffer with 5% BSA) for 1 hour at room temperature (RT). Wells were incubated with 30  $\mu$ L heat-inactivated plasma samples from COVID-19 patients at six serial three-fold dilutions, starting from 1:30 in blocking buffer for 1 hour at RT. The anti-S antibody (abcam, ab273073) anti-N antibody (abcam, ab272852) at nine serial three-fold dilutions, starting from 2  $\mu$ g/mL were used as positive controls. A non-coating well, a non-binding well, and a blank well as negative controls wells were also included on the plate. After washing four times with wash buffer, wells were incubated with peroxidase-conjugated goat anti-human IgG (Sigma, A6029, 1:1,000 dilution), IgA (Sigma, A0295, 1:5000 dilution), or IgM (Sigma, A6907, 1:1000 dilution) antibodies in blocking buffer for 1 hour at RT. Wells were washed four times again before incubating with 30  $\mu$ L 3,3',5,5'-tetramethylbenzidine (TMB) substrate solution (Seracare, 5120-0047). The TMB reaction was stopped after 5 minutes by adding 1M sulfuric acid. The OD at 450nm was measured on a Spectramax Plate Reader. The ELISA antibody titers were defined as the plasma dilutions that result in the middle response of the positive control and calculated by fitting the background-subtracted data to a four-parameter logistic regression model using the R package nplr (55).

## Neutralization assay

The pseudo-virus neutralization assay was conducted by Monogram Biosciences as previously described (56). Briefly, pseudo-typed SARS-CoV-2 virus expressing spike proteins was generated based the original Wuhan-Hu-1 strain sequences (GenBank: NC\_045512.2). Neutralizing antibody titers were measured by incubating nine serial three-fold dilutions of plasma samples with a starting dilution of 1:40 and SARS-CoV-2 pseudo-typed virus at 37°C for 1 hour. HEK-293 cells expressing ACE2 were added to the 96-well plate and incubated for additional 60-80 hours at 37°C for luminescence measurements. Neutralization titers were calculated as the plasma dilution conferring 50% inhibition (ID50) of pseudo-virus infection, adjusting for background luminescence measured from the SARS-CoV-2 nAb positive control.

### Single-cell sequencing data processing

Droplet-based sequencing data were aligned and quantified via Cell Ranger Single-Cell Software Suite (v3.0.0, 10x Genomics) using GRCh38 as a reference. Cells from each demultiplexed sample were first filtered for cells with  $\geq 200$  genes, then filtered based on 1)  $< 10000$  unique molecular identifiers (UMI) counts per cell (library size); 2)  $< 2500$  detected genes per cell; and 3) proportion of mitochondrial gene counts (mitochondrial gene UMIs / total UMIs)  $< 10\%$ . Doublets were simultaneously identified in sample demultiplexing or using scrublet (57) and removed prior to the aforementioned filtering. After QC-based filtering, a total of 846,413 cells were retained for downstream analysis. Scanpy (58) was used to normalize cells via CPM normalization (UMI count per cell was set to  $10^6$ ) and  $\log_{1p}$  transformation (natural log of CPM plus one).

### Single-cell RNA-seq cell type identification

Normalized,  $\ln(\text{CPM}+1)$ , mRNA data from QC-passing single cells were analyzed via PCA (ARPACK). All 50 PCs were used to calculate a neighborhood graph ( $n_{\text{neighbors}}=15$ ) which was utilized to determine UMAP (59) coordinates and Leiden clusters (60). Clusters were assigned cell types based on canonical immune markers and multi-cell-type clusters were separated via additional UMAP and Leiden cluster calculations. Clusters (14,251 cells) that co-expressed markers from multiple cell types were labeled as low-quality or doublets and removed from further analysis. In total, 832,162 cells were deemed high-quality and assigned cell types; these cells did not show noticeable batch differences.

Labeled T cells were used to calculate a  $\text{CD4}^+$  score (sum of min-max-scaled normalized levels of *CD4* transcript and *CD4* surface protein) and a  $\text{CD8}^+$  T cell score (sum of min-max-scaled normalized levels of *CD8A* and *CD8B* transcripts, and *CD8* surface protein). The two scores were min-max-scaled and then projected for manual gating of  $\text{CD4}^+$  and  $\text{CD8}^+$  T cells. T cells with ambiguous scores were classified as “Other T cells”.

### Single-cell regulon-based phenotype identification

Normalized mRNA values for each major immune cell types (B cells,  $\text{CD4}^+$  T cells,  $\text{CD8}^+$  T cells, monocytes and NK cells) were used to construct regulon values via pySCENIC (61). Patient 4 B cells were removed pre-pySCENIC-calculation as they had a cancerous (chronic lymphocytic leukemia) population of B cells. Single cell regulon matrices per cell type were utilized to calculate PCA values (50 PCs). PCs that represent technical factors determined via correlation with n-counts per cell or sequencing batches were removed; specifically, PC1 for B,  $\text{CD4}$ , monocyte and NK populations and PC5 for  $\text{CD8}$ . Remaining PCs were used for neighborhood graph ( $n_{\text{neighbors}}=15$ ) then UMAP and Leiden cluster calculations with resolution input of 0.9, 0.85, 1.2, 0.95 and 1.0 for B,  $\text{CD4}$ ,  $\text{CD8}$ , monocyte and NK, respectively. Cells were then additionally screened for potential doublets; low-quality cells were only detected in B cells where Leiden clusters 7 and 9 were deleted then dimensional reduction and clustering were re-done with PC1 removed due to technical correlation and a Leiden resolution of 0.7. A separate clustering on just memory B cells identified from previous clustering was done to further separate B cell classes using a Leiden resolution of 0.5.

CD4<sup>+</sup> T cells were assigned phenotypes based on canonical marker annotated Leiden clusters with Naïve (Leiden 0); Intermediate (2,3,5,6); Cytotoxic (4) and Hybrid (7). Additional phenotypes T<sub>FH</sub>, Treg and Th<sub>17</sub> were assigned if cells contained normalized mRNA levels above 2.8 (determined via bimodal distribution of mRNA levels from a density plot) for *CXCR5*, *FOXP3*, or *RORC*, respectively, and were not already assigned as a Cytotoxic or Hybrid cell. CD8<sup>+</sup> T cells were similarly labeled with canonical marker described Leiden clusters with Naïve (Leiden 0,10,11,12); Central Memory (6); Effector Memory (4,8); Cytotoxic (1,2,3,5,7,9,13) and Hybrid (14). B cells were annotated as Naïve B cells (Leiden 0), memory B cells (1), intermediate naïve and memory B cells (2), and plasma cells (4). Memory B cells were further annotated with atypical memory B cells (Leiden 2), resting memory B cells (0), and activated memory B cells (1). Monocyte and NK Leiden clusters were annotated based on differential gene expression (Wilcoxon test) and canonical markers. All reduced dimensions (PCA, neighborhood graph, UMAP) and clusters (Leiden) for all of the single cell RNA-seq data were calculated via Scanpy (58).

### Hierarchical clustering of patient trajectories

Transcription factor module values of the single cells for the aforementioned immune classes B, CD4, CD8, monocyte, and NK cells were classified into nine Leiden clusters via a resolution input for Leiden clustering of 0.75, 1.0, 1.0, 0.95 and 1.0, respectively. Percentages of Leiden clusters were calculated per patient timepoint (i.e. blood draw) for each cell type. T2 (acute) and T3 (convalescent) percentages of cell type Leiden clusters were extracted and utilized to cluster patients via Ward's method with criterion "maxclust" and t=4, as visually determined.

### Decision tree analysis

Classification and Regression Trees (CART) recursive partitioning was used to construct a decision tree for identifying patient groups. Independent variables included age, sex, chronic conditions (hypertension, chronic kidney disease, heart failure, chronic kidney disease, diabetes), count of those chronic conditions, medications (immunosuppressants, angiotensin converting enzyme inhibitors, and angiotensin II receptor blockers), minimum blood oxygenation, dyspnea, limitations on activity and disease severity (maximum WOS). For the given set of 15 clinical factors, there were no major observable difference between group 2 and 4 and they were thus combined. Hence, for the CART analysis three patient groups were used as dependent variables (1, 3, 2 and 4). Ten-fold cross-validation was applied to evaluate tree reliability, and gini index was used to quantify the impurity and select splits during classification. To avoid overfitting, a maximum acceptable difference in risk between the pruned and the sub-tree was set at one standard error. Missing data were handled by surrogate splits. The following formula were used to calculate misclassification rate or prediction error in cross-validation (root node error x relative error x 100) and prediction accuracy (1 – misclassification rate).

### Plasma proteomic Gene Ontology (GO) analysis

Batch-corrected plasma protein levels were converted into Z-scores using the means and the standard deviations estimated for the residuals in the matched control samples, which included

corrections for age, sex, and body mass index (BMI). The Z-scores thereby indicate the difference in plasma abundance relative to the control sample - e.g., a Z-score of 2 represents 2 standard deviations above the expected control value. The Z-scores were used for plotting the line plot in Fig. 1e middle panel. A score for a GO process was calculated by first summing the averaged Z-scores from GO-affiliated plasma proteins with a Z score larger than 2 or smaller than -2. Those proteins were termed significant outliers. The resulting sum was multiplied by the fraction of significant outliers relative to all possible measured proteins associated with that GO process to yield a GO score. For example, for Developmental Processes, Group 4 patients, T3 time point: 64 proteins are positive outliers, and one protein is a negative outlier, with Z-scores summing to 192.9. There were 243 proteins measured that were associated with this GO process, so  $192.9 \times 65 / 243 = \text{GO Score of } 51.6$ . Such scores were used for plotting Fig. 1d bottom panel.

### Clinical factor and plasma analyte correlation analysis

We estimated Pearson correlations between clinical information and EHR data (clinical factors) and plasma abundance of proteins and metabolites. Clinical information and lab tests were extracted from EHRs; continuous variables were directly utilized, and categorical variables were binarized when possible. Multi-class categorical variables were binarized by performing separate comparison between each non-standard category with the standard. For example, the multi-class variable “hypoxia” contained categories resolved, unresolved and no hypoxia leading to the comparisons resolved vs. no hypoxia, unresolved vs. no hypoxia, and had hypoxia (unresolved or resolved) vs. no hypoxia. In all cases no hypoxia was set to 0 and the other category was set to 1. All analyte values were from acute (T2) measurements.

Correlations that had  $|r| \geq 0.5$  (magnitude of r),  $p\text{-value} < 0.05$  and  $n\text{-samples} > 15$  (number of blood draws used for correlation) were deemed significant. Remaining x (plasma analyte) and y (clinical factors) were further filtered via their number of significant correlations. Lists of x and y analytes were progressively filtered by removing analytes until all analytes had at least three significant correlations with others. Resulting correlation heatmap was ordered by non-filtered correlation values and heatmap was masked by only showing the significant correlations. Clustering was done via Ward’s method with criterion “maxcluster”, t “8” (“5”) for plasma (clinical factors), as visually determined.

Protein scores were calculated via standard scaling (subtract mean, divide by standard deviation) the proteins within defined plasma analyte clusters then averaging the scaled values per patient per plasma analyte cluster.

### Differential metabolites across patient groups

Metabolites with all NaN values were removed and only values from healthy donors or convalescent draws of patients from patient groups were kept. Differential expression of metabolites was computed via `scanpy.tl.rank_genes_groups` (method=“Wilcoxon”, n\_genes=all variables). Each metabolite was assigned to the group with the highest differential score; the second-highest differential score was subtracted from the highest to give a normalized differential score which ranked metabolites



intra-group, the top30 metabolites per group were taken. Intra-group ordering of metabolites were done via the same methods as the previous correlation analysis.

### Inter-omic correlation analysis

Analytes and values included in Spearman correlation analysis are provided in table S10. Significant correlations were defined, and ordering and clustering performed in the same manner as the previous correlation analysis with  $t = 14$ . The top5 analytes (ranked via n-significant-correlations) from plasma clusters, 10x-derived phenotypes, clinical information, lab tests and top25 analytes from functional data were selected.

### Single-cell TCR-seq data processing

Droplet-based sequencing data for T cell receptor sequences were aligned and quantified using the Cell Ranger Single-Cell Software Suite (version 3.1.0, 10x Genomics) against the GRCh38 human VDJ reference genome.

### Single-cell TCR phenotype associations

Filtered annotated contigs for TCRs were analyzed via scirpy (62). Aforementioned contigs were filtered for either CD4<sup>+</sup> or CD8<sup>+</sup> T cells (as identified via single cell RNA-seq analysis) and then subject to clonotype definition and clonal expansion analysis utilizing nucleotide sequences. Samples were then concatenated together and merged with gene expression data for simultaneous single cell TCR and RNA data visualization.

Both the integrated CD4<sup>+</sup> and CD8<sup>+</sup> T cell datasets were subject to filtering for cells with complete TCR sequences, defined as a detectable TRA and TRB. TCRs were normalized per sample (patient blood draw) by sampling with (without) replacement TCRs of samples with n-TCRs  $< (\geq)$  median TCRs per sample. Pheno-tags were created by compounding cell phenotype with blood draw timepoint (filtered for acute and convalescent). TCR x pheno-tag matrix was constructed with values as the percent of cells in the given pheno-tag with the given TCR. Only TCRs present in  $\geq 2$  pheno-tags were included, and values were normalized to  $\ln(\text{value}+1)$ . The matrix was then ordered and clustered in the same manner as the correlation analyses with  $t$  set to "5", as visually ascertained.

### Differential gene expression and pathway enrichment analysis of TCR clonal trajectories

For CD8<sup>+</sup> T cells, TCR group 1 and 5 were considered clonally deleted and expanded respectively. Differential analysis was performed via scanpy.tl.rank\_genes\_groups (method="Wilcoxon", n\_genes=300) on single cells comparing clonally deleted vs. expanded cells. For CD4<sup>+</sup> T cells, TCRs were hyper-clustered with  $t = 14$  and cluster 5 and 10 were considered clonally deleted and expanded respectively. Pseudo-bulk values per clonal-trajectory (averaged across single cells) were utilized in GSEA analysis to derive pathway scores.

## GLIPH2 Analysis

The TCRs of interest were run with a database to cluster TCRs with TCRs of known specificity. Two distinct GLIPH2 were run. For CD8<sup>+</sup> T cells, we utilized a database from Adaptive Immunity of TCRs matched to SARS-CoV-2 epitopes (63). For CD4<sup>+</sup> T cells, we utilized a database of CD4<sup>+</sup> T cells that were reactive to SARS-CoV-2 peptides. In both cases, GLIPH2 was run without a reference. The output was modified to only include clusters that shared a V gene. The output was visualized as a binary heat map where positive=at least 1 TCR from the cluster occurred in a GLIPH2 group with the respective database TCR.

## Single-cell BCR & RNA-seq integration

Annotations from sc-RNAseq were used to define memory B cells in the sc-BCR data. Somatic hypermutation rates (SHM) were defined as the percentages of gaps and mismatches in the query sequence compared to the top germline V gene hit identified through IgBLAST (64) for each contig. Filtered contig outputs from the 10x Genomics Cell Ranger pipeline were used as input to the R package Immunarch (65) to assign clonotypes to memory B cells for each T3 blood draw for calculation of isotype usage in Fig. 3g.

## BCR convergent sequence analysis

IGH sequences from single cells with paired productive heavy and light chains were searched against a list of known SARS-CoV-2 binding antibodies to identify convergent sequences according to the following criteria: utilization of the same IGHV and IGHJ genes; same CDR-H3 lengths; and CDR-H3 amino acid sequences that were within a Hamming distance cutoff of 15% of the length of the CDR-H3.

## Machine learning

Z-scores of plasma protein abundance at baseline (T1) were used to construct a binary classifier to predict patient group status (including death). Analytes were pre-filtered for Z-score $\geq$ 2 in any patient group or Z-score-difference $\geq$ 2 between any patient group pair. For a given random state, the top15 most important features derived from ExtraTreesClassifier (n\_estimators=100) per pair were concatenated into a list of top features. Top15 most represented features from aforementioned list was used to derive all possible five marker combinations. Each combination was tested via StratifiedShuffleSplit cross validation (CV) with train,test=2/3,1/3 percentages. For each train,test combination GridSearchCV (scoring="accuracy") was utilized with logistic regression and range of C's from 10<sup>-2</sup> to 10<sup>13</sup> using StratifiedShuffleSplit in the same manner as before, AUC scores were recorded via sklearn's roc\_curve and auc methods.

Models (marker combination and random state) were then filtered by the average minimum AUC of pair-wise comparisons across CVs and ranked by the number of CVs that have a minimum AUC > 0.7 resulting in the model with random state 2 using markers FIGF, KRT19, ENTPD5, LILRA5 and AREG.

Marker significance was then approximated via its coefficient in the logistic regression model corresponding to a given pair-wise comparison. Mean AUC scores were obtained by averaging AUCs across CV iterations and ROC curves were plotted via sklearn's `plot_roc_curve` per CV iteration and binary classification.

### Statistical analysis

Statistical analyses were performed using Python or R software packages. Null hypotheses between two groups used in all bar plots and box plots were tested using the non-parametric Mann-Whitney U test. Other specific statistical tests and their significance levels were denoted in each figure legend.

**Acknowledgments:** We appreciate the insightful discussion from Prof. David Baltimore, Prof. David Koelle, Prof. Alan Aderem and the ISB COVID-19 Study Group. We are grateful to all participants in this study and to the medical teams at Swedish Medical Center for their support. We thank the Northwest Genomic Center for help with sequencing services, and the ISB-Swedish COVID-19 Biobanking Unit. We thank Amazon Web Services for their support through cloud computing credits provided by the AWS Diagnostic Development Initiative (DDI). We acknowledge funding support from the Wilke Family Foundation (JRH), the Murdock Trust (JRH), Gilead Sciences (JRH), the Swedish Medical Center Foundation (JDG), the Parker Institute for Cancer Immunotherapy (JRH, MMD, PDG, LLL, AR and JAB), Merck, and the Biomedical Advanced Research and Development Authority (HHSO10201600031C to JRH). KW was funded by DOD (W911NF-17-2-0086), NIH (R01 DA040395 and UG3TR002884). RG was funded by the NIH Human Immunology Project Consortium (U19AI128914) and the Vaccine and Immunology Statistical Center (Bill and Melinda Gates Foundation OPP1032317). Further funding by NIH (AI068129 to LLL and R21 AI138258 to NS). YS was supported by the Mahan Fellowship at Herbold Computational Biology Program of Fred Hutch Cancer Research Center.

### **Author contributions:**

Conceptualization: YS, JDG, JRH; Methodology: YS, DY, DGC, JRH; Formal Analysis: YS, DGC, DY, AJPB, SAK, RG, JRH; Investigation: YS, DY, DGC, JC, CLD, SH, RZ, JX, SL, KS, AJPB, SD, CL, RHN, SAK, PB, VRD, KGA, JL, FY, CR, PT, BS, JZ, SM, KM, RE, LJ, YZ, LR, RL, WC, DSOM, CRD, JAW, HAA, WW, NDP, NS, JEH, ATM, KW, AR, LLL, SDB, JAB, LH, RG, PDG, MMD, JDG, JRH; Resources: JDG, JRH; Writing – Original Draft: YS, DY, DGC, JRH; Writing – Review & Editing: YS, DY, DGC, JC, CLD, SH, RZ, JX, SL, KS, AJPB, SD, CL, RHN, SAK, PB, VRD, KGA, JL, FY, CR, PT, BS, JZ, SM, KM, RE, LJ, YZ, LR, RL, WC, DSOM, CRD, JAW, HAA, WW, NDP, NS, JEH, ATM, KW, AR, LLL, SDB, JAB, LH, RG, PDG, MMD, JDG, JRH.

**Competing interests:** JRH and AR are founders and board members of Isoplexis and PACT Pharma. MMD is a member of the Scientific Advisory Board of PACT Pharma. JAB is a member of the Scientific Advisory Boards of Arcus, Solid and VIR. JAB is a member of the Board of Directors of Gilead and Provention. JAB is the CEO of Sonoma Biotherapeutics. LLL is on the scientific advisory boards of Alecto, Atreca, Dragonfly, DrenBio, Nkarta, Obsidian Therapeutics, SBI Biotech. RG has received consulting income from Juno Therapeutics, Takeda, Infotech Soft, Celgene, Merck and has received research support from Janssen Pharmaceuticals and Juno Therapeutics, and declares ownership in CellSpace Biosciences. PDG

is on the Scientific Advisory Board of Celsius, Earli, Elpiscience, Immunoscope, Rapt, and Nextech, was a scientific founder of Juno Therapeutics, and receives research support from Lonza. JDG declared contracted research with Gilead, Lilly, and Regeneron. The remaining authors declare no competing interests.

**Fig. 1. Overview of the Longitudinal Multi-Omic Analysis of Immune Responses in COVID-19 Patients.**

(A) Overview of ISB/Swedish INCOV Study design. The various analytic assays run on plasma and isolated PBMCs are shown. Detailed patient information listed in table S1. (B) Hierarchical clustering of patients based on immune cell subtypes at T2 and T3 resolved four patient-group trajectories. Each column (row) corresponds to the percentage of an immune subtype (patient) at acute (T2) or convalescence (T3). Top three rows indicate patient sex, age and patient group. Left two columns indicate major immune class and measurement time. Heatmap keys are provided beneath or to the right of the heatmap. Full list of subpopulations and their percentages provided in table S2. (C) Decision tree classification model showing the top variables that discriminate patient groups using demographic and clinical variables recorded at T1 and T2. Colors of boxes the majority patient group (see color key) for the Leaf-node. Percentages within boxes indicate prediction accuracy of patient groups for that leaf node. Patient numbers per leaf node are listed in table S3. (D) RNAemia viral load analysis. Boxplot showing quantified plasma viral loads for samples of different severities at diagnosis (T1) and acute (T2). Y-axis unit is 40-CT (cycle threshold). Data points were color coded by patient groups. Data points from patients with no group assignments due to unavailable T3 blood draws are shown in grey. This data is listed in table S4. (E) Plasma proteomic analysis across patient groups. Top (bar plot), disease severity of each group per time point, plotted as mean±SEM. Middle (line plot), plasma abundance of five severity-related proteins (quantified by Z-scores, Supplemental Methods) per patient group. From left to right, each group of lines shows data at T1, T2, and T3, plotted as mean±SEM. Bottom, gene ontology (GO) score (Supplemental Methods) of three biological processes (infection responses) in each patient group. The three groups of bars for each patient group represent data at T1, T2, and T3. (F) Inter-omic correlations between plasma analytes and clinical data. Each row is a selected clinical measurement from electronic health records and each column is a plasma protein or metabolite. Top correlative clinical metrics are shown, full map shown in Fig. S1D. Color represents Pearson correlation coefficient of row and column variables (see color key at bottom). Significant correlations ( $|r| \geq 0.5$  and  $p\text{-value} < 0.05$ , Supplemental Methods) are shown. (G) Levels of plasma proteins in cluster 5 and 7 across patient groups at T3. Boxplot center, median; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range (IQR). \* $p\text{-value} < 0.05$ , \*\* $p\text{-value} < 0.01$ , \*\*\* $p\text{-value} < 0.001$ , \*\*\*\* $p\text{-value} < 0.0001$ .

**Fig. 2. Phenotypical, functional and TCR analysis of CD8+ and CD4+ T cells in patients from the four recovery trajectories.**

(A) UMAP embedding of CD8+ (top) and CD4+ T cells (bottom) colored by unsupervised clustering. (B) Box plots showing percentages of CD8+ and CD4+ T cell phenotypes in each patient group at acute (T2) (top) and convalescent (T3) (bottom). (C) Single-cell polyfunctional strength index (PSI) of CD8+ (top) and CD4+ (bottom) T cells in each patient group, plotted as mean±SEM. (D) Hierarchical clustering of CD8+ (left) and CD4+ (right) T cell TCRs (rows) based on TCR sharing patterns across select phenotypes and time points (see color key at bottom). Only cytotoxic, effector-memory or intermediate phenotypes are shown. The full heatmap is provided in Fig. S4A. (E) TCR-transcriptome lineage tracking analysis for

CD8+ (CD4+) T cells. Top, cartoon illustration for analyzing differential gene signatures for cytotoxic T cells at T2 that will convert into memory (expanded) at T3 versus those will clonally contract at T3. Bottom left, top genes differentially expressed between the two subsets of cytotoxic CD8+ T cells in T2. Bottom right, top pathways differentially regulated at T2 for cytotoxic CD4+ T cells that clonally contracted versus clonally expanded. (F) Representative GLIPH2 groups of TCRs from TCR clusters 3 and 4 (defined in (D), left) for CD8+ TCRs (left two panels) and cluster 2 (defined in (D), right) for CD4+ TCRs (right panel). Blue represents published SARS-CoV-2-specific TCRs, tan represents TCRs from our dataset. Inferred epitopes or virus proteins based on the blue TCR are provided on the right. Boxplot center, median; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range (IQR). \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

### Fig. 3. Longitudinal tracking of B cell responses in COVID-19.

(A-B), Box plots showing IgG and IgA antibody titers at T3 measured by ELISA to RBD (A) and spike protein (B). (C) Heatmap with hierarchical clustering showing normalized IgG, IgA and IgM antibody titers to RBD, spike and nucleocapsid protein. (D) Pseudovirus neutralization assay of plasma. Left, diagram illustrating the neutralizing assay. Middle, box plot showing neutralizing antibody titers in each patient group at T2 and T3. Samples from the same patient are connected with a red line. Right, linear regression plot with 95% confidence level (grey shaded areas) showing the relationship between neutralizing antibody titers and anti-RBD IgG antibody titers at T3. Pearson correlation coefficient is shown. Points were color-coded by patient group (see color key at top-right). (E) Box plot showing percentages of antibody-secreting cells (ASC) in healthy individuals and each patient group across all time points. (F) Box plot showing somatic hypermutation (SHM) rates in memory B cells across all time points. (G) Box plots showing percentages of IGHG1 (left) and IGHM (right) memory B cells over all memory B cells at T3. (H) Convergent BCR CDR-H3 sequences that are specific for SARS-CoV-2 spike protein shared within some patients in our patient cohort. Boxplot center, median; box limits, 25th and 75th percentiles; whiskers, 1.5× interquartile range (IQR). \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

### Fig. 4. Integrated analysis of functional, phenotypical, clinical data across patients and machine learning model for predicting patient groups at diagnosis.

(A) Correlation matrix of 45 functional, phenotypical, plasma and clinical features from COVID-19 patients. Color indicates the Spearman correlation coefficient between row and column variables. Average levels of each measurement for each patient group are shown as the top heatmap. Datasets used for generating this correlation matrix are listed in table S10. (B) Machine learning model for predicting patient groups at diagnosis using five biomarkers. Left, cartoon illustration of machine learning model. Middle, classification accuracy evaluated by AUC scores for pair-wise classification between patient groups using five markers measured at diagnosis. Right, heatmap depicting feature importance of each marker in pair-wise comparisons.

### Fig. S1. Overview of the Longitudinal Multi-Omic Analysis of Immune Responses in COVID-19 Patients

- (A) Cartoon illustration of patient recovery trajectory calculation based on immune cell subpopulation percentages from early infection (T2) to recovery (T3).
- (B) Bar plot showing age, viral load or clinical metrics across patient groups at acute (T2) or convalescence (T3). P-values were FDR-corrected.
- (C) Analysis of group-unique metabolites at T3 for patient groups and healthy unexposed individuals. Top, heatmap of top30 metabolites uniquely upregulated per group. Columns indicate metabolites and rows indicate patient groups at T3 or unexposed healthy donors (see color key at bottom). Bottom, box plots showing plasma abundance of selected metabolites per group and levels across four patient groups at T3 or of healthy donors.
- (D) Inter-omic correlations between plasma-omic analytes and clinical data (full version of Fig. 1f). Left, heatmap of cross-omic correlations. Each row (column) corresponds to a clinical measurement from electronic health records (plasma protein or metabolite). Color represents Pearson correlation coefficient of row and column variable (see color key at bottom).  
Boxplot center, median; box limits, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× interquartile range (IQR). Right, abundance of plasma analyte cluster across patient groups at T3. \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

**Fig. S2. Phenotypical and functional analysis of CD8<sup>+</sup> T cells in patients with four different recovery trajectories.**

- (A-C) UMAP embedding of CD8<sup>+</sup> T cells colored by unsupervised clustering (left panel in **A**), levels of selected mRNA transcripts (other panels in **A**), surface proteins (**B**) and clonal size (**C**).
- (D) Clonal expansion score for CD8<sup>+</sup> T cells of samples from patient groups and healthy unexposed individuals at acute (T2) (top) and convalescent (T3) (bottom).
- (E) UMAP embedding density of CD8<sup>+</sup> T cells per patient group and time point (T2 and T3). Selected clusters are encircled (see color key in **A**).
- (F) Box plots showing CD8<sup>+</sup> T cell phenotype percentages at acute (T2) (top) and convalescent (T3) (bottom) in each patient group.
- (G) Box plots indicate percentage of CD8<sup>+</sup> T cells secreting selected cytokine/chemokines for samples from different patient groups at convalescence or from unexposed healthy donors.
- (H) Heatmap visualization of average cytokine secretion frequency for cells from patient groups at convalescence or healthy unexposed individuals.

Boxplot center, median; box limits, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× interquartile range (IQR). Right, abundance of plasma analyte cluster across patient groups at T3. \*p<0.05, \*\*p<0.01, \*\*\*p<0.001, \*\*\*\*p<0.0001.

**Fig. S3. Phenotypical and functional analysis of CD4<sup>+</sup> T cells in patients from the four recovery trajectories.**

- (A-C) UMAP embedding of CD4<sup>+</sup> T cells colored by unsupervised clustering (left panel of **A**), levels of selected surface proteins (other panels of **A**), mRNA transcripts (**B**), and clonal size (**C**).

- (D) Clonal expansion score for CD4<sup>+</sup> T cells of samples from patient groups and healthy unexposed individuals at acute (top) and convalescent (bottom).
- (E) UMAP embedding density of CD4<sup>+</sup> T cells for samples across patient groups and time (T2 and T3). Selected clusters are encircled (see color key in A).
- (F) Bar plot of clonal expansion levels per CD4<sup>+</sup> T cell phenotype color-coded by clonal expansion sizes (n=1, n=2-4, n≥5).
- (G) Box plots showing the percentages of CD4<sup>+</sup> T cell phenotypes at acute (T2) (top) and convalescent (T3) (bottom) stages for each patient group.
- (H) Box plots indicate percentages of CD4<sup>+</sup> T cells secreting selected cytokine/chemokines for samples across patient groups at convalescence or from healthy unexposed donors.

Boxplot center, median; box limits, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× interquartile range (IQR). \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

**Fig. S4. Lineage tracking reveals TCR clusters of distinct clonal dynamics and early signatures for memory T cells formation versus clonal contraction.**

- (A) Hierarchical clustering of CD8<sup>+</sup> (left) and CD4<sup>+</sup> (right) TCRs (rows) based on TCR sharing patterns across phenotypes and time (columns, see color key on right of heatmaps). Full heatmap with all T cell phenotypes for Fig. 2g.
- (B) TCR clonal shifting patterns from T2 to T3 across patient groups. Box plots showing the percentages of TCRs from each TCR cluster (defined in A) over all TCRs for each patient group at acute (T2) and convalescent (T3) time points, for CD8<sup>+</sup> (left two panels) and CD4<sup>+</sup> (right two panels) T cells.
- (C) Hierarchical clustering of inferred SARS-CoV-2 epitopes for each TCR from CD8<sup>+</sup> (left) and CD4<sup>+</sup> (right) T cells. Rows, TCR clusters from A. Columns, inferred epitopes or viral proteins specific for each TCR (left) or reference SARS-CoV-2 specific CD4<sup>+</sup> TCR (right). Beige color indicates given TCR cluster contains column's viral epitope/protein.

Boxplot center, median; box limits, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× interquartile range (IQR). \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

**Fig. S5. B cell responses in COVID-19**

- (A-C) UMAP embedding of B (top) and memory B cells (bottom) color-coded by unsupervised clustering A, levels of mRNA transcripts (B) and surface proteins (C).
- (D-E) Heatmap showing normalized expression of selected mRNA transcripts (D) and surface proteins (E) per cluster (defined in A) in B (top) and memory B cells (bottom).
- (F) Box plots showing somatic hypermutation rates in Ig gamma heavy chains (left), mu heavy chains (middle) and light chains (right) of memory B cells at T3.
- (G) Box plots showing IgM antibody titers against RBD, spike and nucleocapsid protein, and IgG and IgA antibody titers against the nucleocapsid protein measured at T3.

Boxplot center, median; box limits, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× interquartile range (IQR). \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

**Fig. S6. Phenotypical and functional analysis of monocytes.**

- (A) UMAP embedding of monocytes colored by unsupervised clustering (left panel) and levels of selected mRNA transcripts and surface proteins (other panels).
- (B) UMAP embedding density of monocytes for samples from different patient groups at acute (T2) and convalescent (T3) time points. Selected clusters are encircled (see color key in A).
- (C) Box plots showing the percentage of non-classical monocytes and myeloid-derived suppressor cells (MDSCs) at acute (T2) and convalescent (T3) stages for each patient group. Data are represented as mean±SEM.
- (D) Single-cell polyfunctional strength index (PSI) of monocytes in each patient group. Data are represented as mean±SEM.
- (E) Heatmap visualization of average cytokine secretion frequency for monocytes for each patient group at convalescence or healthy unexposed individuals.

\*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

**Fig. S7. Phenotypical and functional analysis of NK cells.**

- (A-B) UMAP embedding of NK cells colored by unsupervised clustering (left panel of A) and levels of selected surface proteins (other panels of A) and mRNA transcripts (B).
- (C) UMAP embedding density of NK cells for samples in each patient group and time points (T2 and T3). Selected clusters are encircled (see color key in a).
- (D) Box plots showing the percentages of NK cell phenotypes at acute (T2) and convalescent (T3) stages for each patient group.
- (E) Bar plot showing *GZMH* transcript levels (left) and NK cell effector PSI (right) in each patient group and healthy unexposed individuals. Data are represented as mean±SEM.
- (F) Heatmap visualization of average cytokine secretion frequency for NK cells for each patient group at convalescence or healthy unexposed individuals.

Boxplot center, median; box limits, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× interquartile range (IQR). \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

**Fig. S8. Correlations between different measurements**

Scatter plots depicting correlation among different measurements across patients with 95% confidence intervals shaded in blue. Spearman's correlation coefficients and p-values are shown. \*p-value<0.05, \*\*p-value<0.01, \*\*\*p-value<0.001, \*\*\*\*p-value<0.0001.

**Fig. S9. Machine learning model evaluation on validation sets and levels of five biomarkers at T1 across patients.**



**(A)** Receiver operating characteristic curves, per cross-validation (CV) iteration, for pair-wise classification (see subtitles) based on the levels of 5 markers at T1 for different validation pairs. Area-under-curve (AUC) values for different CVs (in different colors) are displayed.

**(B)** Boxplots showing plasma abundance of the five plasma biomarkers levels across the four patient groups or deceased patients at T1.

Boxplot center, median; box limits, 25<sup>th</sup> and 75<sup>th</sup> percentiles; whiskers, 1.5× interquartile range (IQR).

**Fig. S10. Summary of immune features across four convalescent patient groups and healthy donor.**

**Table S1. Clinical data for each patient and healthy donor**

Clinical measurements and observations of each blood draw of COVID-19 patients and healthy donors included in our study are shown. The Study Subject IDs in each of the four immunological trajectory groups identified in our study are included as well.

**Table S2. Immune cell subpopulation percentages**

The percentages of each immune subpopulation of B cells, CD4<sup>+</sup> T cells, CD8<sup>+</sup> T cells, monocytes and NK cells for each sample are shown.

**Table S3. Patient number in each leaf node**

Numbers of patients of each group in each leaf node in **Fig. 1C** are shown. The group with majority of patients and percentages of the major patient group in that leaf node are shown.

**Table S4. Viral load of plasma and nasal swabs**

Cycle threshold (CT) values of plasma viral load (RNAemia) and nasal swab viral load of each sample are shown. A CT value > 40 is labeled with “ND” and a missing value due to the unavailability of the sample is labeled with “Not available”.

**Table S5. Plasma proteomics and metabolomics data**

The batch-corrected proteomics and metabolomics data adjusted for age, sex, and BMI are included. The proteomics data further converted to Z-scores are also included.

**Table S6. Top 30 upregulated metabolites in each patient group**

The top 30 upregulated metabolites per group calculated through differential expression are shown.

#### **Table S7. Plasma analyte clusters**

Proteins and metabolites in each of the nine plasma analyte clusters in **Fig. 1F** are shown.

#### **Table S8. TCR clusters**

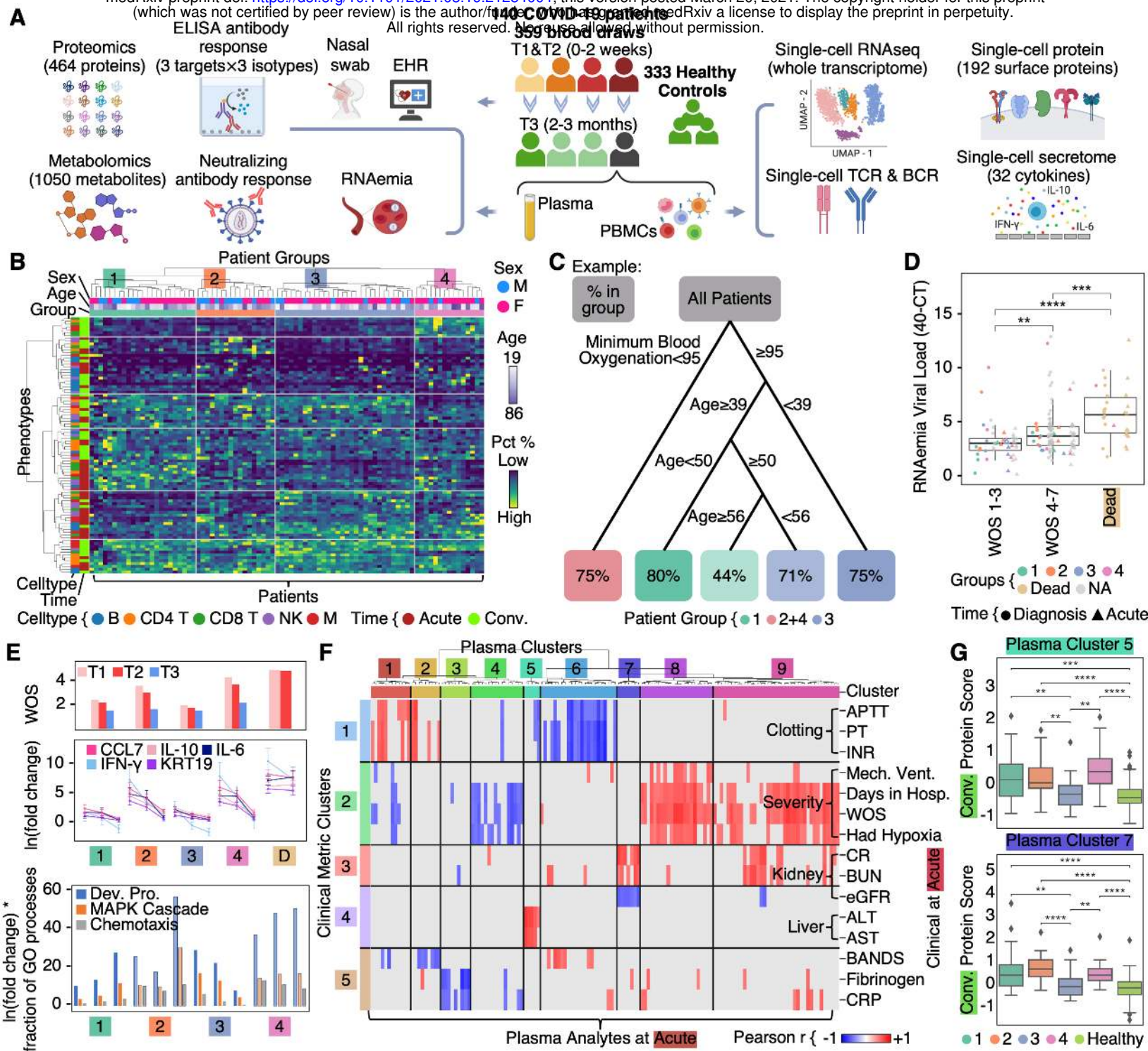
The CDR3 nucleotide sequences of paired TCR $\alpha$  and TCR $\beta$  chains in each TCR cluster in **Fig. 2D**. CD8 and CD4 TCR Clusters are shown separately.

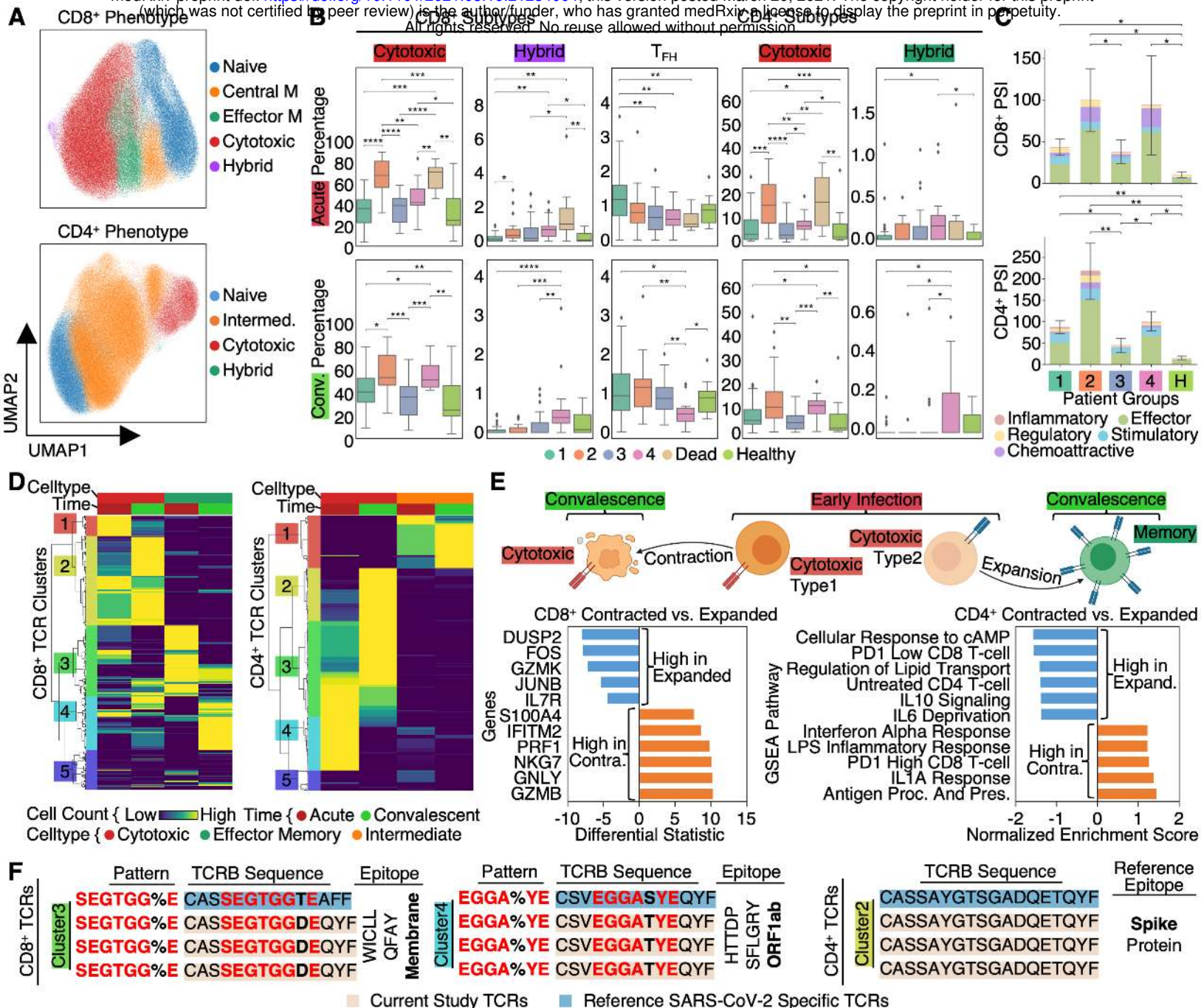
#### **Table S9. SARS-COV-2 epitopes identified via GLIPH2 analysis for TCR clusters**

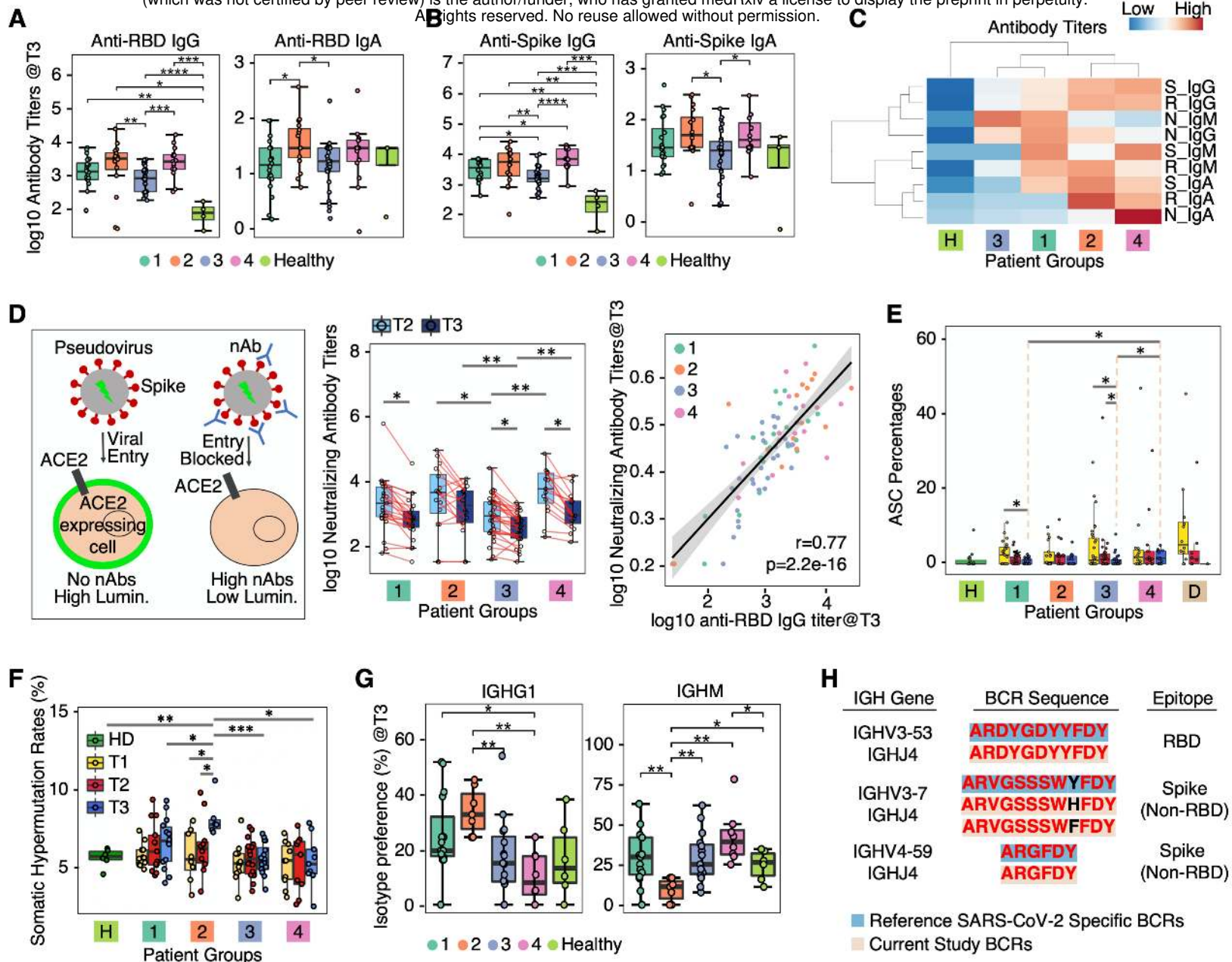
The amino acid sequences of SARS-COV-2 epitopes that the TCR clusters in **Fig. 2D** have a specificity for are shown. A “1” for a cluster indicates a TCR specific for that epitope could be identified in this cluster.

#### **Table S10. Datasets used for generating the correlation matrix in Fig. 4A**

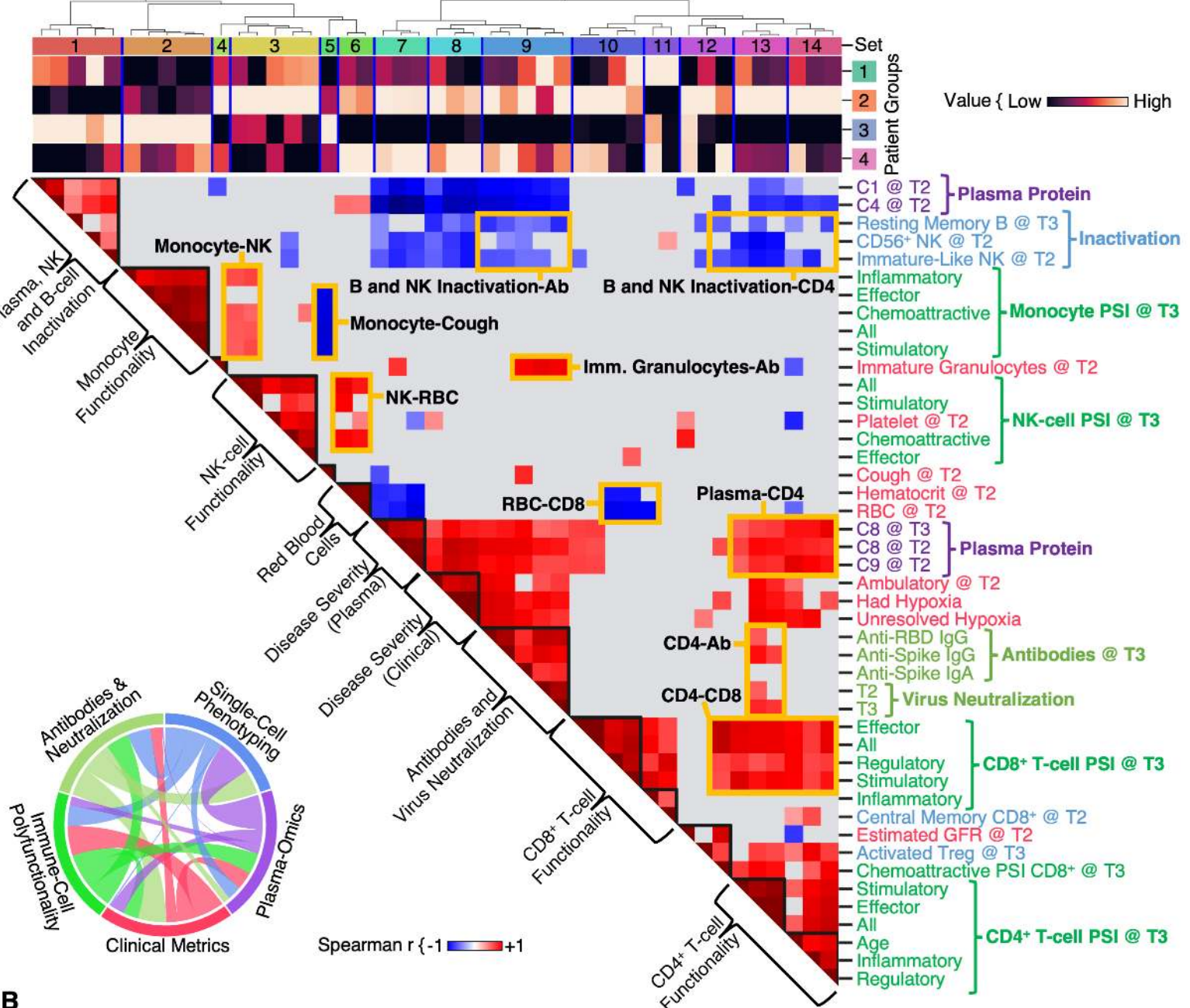
282 features derived from multiple datasets in our study for each patient are shown. Cell type, assay type, and the time of the blood draw are indicated as well.







**A**



**B**

