

Longitudinal single-cell analysis of SARS-CoV-2-reactive B cells uncovers persistence of early-formed, antigen specific clones

Lydia Scharf¹, Hannes Axelsson¹, Aikaterini Emmanouilidi¹, Nimitha R. Mathew¹, Daniel J Sheward², Susannah Leach^{1,3}, Pauline Isakson⁴, Ilya V. Smirnov¹, Emelie Marklund^{5,6}, Nicolae Miron⁴, Lars-Magnus Andersson^{5,6}, Magnus Gisslén^{5,6}, Ben Murrell², Anna Lundgren^{1,4}, Mats Bemark^{1,4,*}, Davide Angeletti^{1,*}

¹Department of Microbiology and Immunology, Institute of Biomedicine, University of Gothenburg, Gothenburg, Sweden

²Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden

³Department of Clinical Pharmacology, Sahlgrenska University Hospital, Gothenburg, Sweden

⁴Department of Clinical Immunology and Transfusion Medicine, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden

⁵Department of Infectious Diseases, Sahlgrenska University Hospital, Region Västra Götaland, Gothenburg, Sweden

⁶Department of Infectious Diseases, Institute of Biomedicine, University of Gothenburg, Gothenburg, Sweden

LS, HA and AE contributed equally to the study

*Correspondence:

Davide Angeletti, Department of Microbiology and Immunology, Institute of Biomedicine, Medicinaregatan 7A, 413 90 Gothenburg, Sweden. Phone: +46-(0)31-786 6324. Email: davide.angeletti@gu.se

Or

Mats Bemark, Department of Microbiology and Immunology, Institute of Biomedicine, Medicinaregatan 7A, 413 90 Gothenburg, Sweden. Phone: +46-(0)70-323 7043. Email: mats.bemark@immuno.gu.se

The authors have declared that no conflict of interest exists.

ABSTRACT

Understanding persistence and evolution of B cell clones after COVID-19 infection and vaccination is crucial for predicting responses against emerging viral variants and optimizing vaccines. Here, we collected longitudinal samples from severe COVID-19 patients every third to seventh day during hospitalization and every third month after recovery. We profiled their antigen-specific immune cell dynamics by combining single cell RNA-Seq, Cellular Indexing of Transcriptomes and Epitopes by Sequencing (CITE)-Seq, B cell receptor (BCR)-Seq with oligo-tagged antigen baits. While the proportion of Spike Receptor Binding Domain-specific memory B cells (MBC) increased from 3 months after infection, the other Spike- and Nucleocapsid-specific B cells remained constant. All patients showed ongoing class switching and sustained affinity maturation of antigen specific cells, which was not significantly increased early after vaccine. B cell analysis revealed a polyclonal response with limited clonal expansion; nevertheless, some clones detected during hospitalization, as plasmablasts, persisted for up to one year, as MBC. Monoclonal antibodies derived from persistent B cell families increased their binding and neutralization breadth and started recognizing viral variants by 3 months after infection. Overall, our findings provide important insights into the clonal evolution and dynamics of antigen specific B cell responses in longitudinally sampled COVID-19 infected patients.

INTRODUCTION

Since the emergence of SARS-CoV-2 in December 2019, there have been over 630 million cases and at least 6.5 million deaths worldwide (1, 2). Despite an extensive vaccination campaign, which substantially reduced morbidity and mortality, the virus is still in circulation mainly due to the appearance of viral variants which escape pre-existing immunity.

Several research groups have described the early immune response upon infection (3-6). During severe infection, general lymphopenia is associated with an increased number of circulating plasmablasts (3), Th1-like CD8 and CD4 cells (5), megakaryocytes and erythroid cells (7). In serum, Spike (S)-binding, neutralizing antibodies (Abs) of the IgA and IgG isotypes emerge early after COVID-19 infection, even before IgM, as reported in some studies (8). Further, it has been suggested that the early plasmablast burst originates from the reactivation of memory B cells (MBC), specific for seasonal beta-coronaviruses (i.e. HKU1 and OC43) (9-14).

With the emergence of viral variants, there has been great emphasis in studying MBC. Early studies with influenza, dengue and other viral infections in animal models suggested that the MBC pool has greater breadth of antigenic binding, as compared to the plasmablast response (15-18). This led to the hypothesis that while plasma cells and the serum Abs they produce protect against reinfection with the same strain, the MBC pool represents a diverse reservoir which is able to protect against possible emerging variants. Several studies have now followed MBC development after SARS-CoV-2 infection and reported a continuous increase of B cell receptor (BCR) mutations, consistent with antigen persistence and ongoing germinal center (GC) activity (12, 19-23). The increased number of mutations was also linked with increased affinity and, importantly, neutralization breadth. Interestingly, during influenza infection in animal models, GC persistence has been observed for over 180 days, suggesting this to be a common feature of acute viral infections (24, 25).

Monoclonal antibodies (mAbs) cloned from COVID-19 patients at different time points after infection demonstrated increased neutralizing breadth against viral variants, even from mAbs belonging to the same clonal family (21, 22). Other studies investigated BCR characteristics during disease (26, 27). Finally, work from the Wilson's lab linked transcriptional program of

single B cell with VDJ properties and antigen specificity, in a cross-sectional cohort (9). However, no study followed the same patients longitudinally during hospitalization, after recovery and upon vaccination to investigate immune responses, BCR characteristics and antigen specificity.

To address this, six COVID-19 patients were recruited at hospital admission and followed during disease and after recovery, for up to one year. Half of the patients were also vaccinated by the last time point. We analyzed total peripheral blood mononuclear cells (PBMCs) and B cells using single cell transcriptomics, expression of 138 surface proteins, antigen binding (S, Receptor Binding Domain (RBD) or Nucleocapsid (N)) and BCR sequences at each of the analyzed time points. Our longitudinal approach was able to deduct the origin of antigen-specific B cells and their evolution within each patient. Further, by expressing persisting clones as mAbs, we demonstrate that such clones can be detected within three days after hospital admission, persist up to one year and progressively increase their neutralization breadth. Overall, our longitudinal study provides important insights into B cell evolution after viral infections.

RESULTS

Longitudinal analysis of peripheral immune responses in COVID-19 patients

To better understand the temporal dynamics of the immune response within individual patients, we focused on six patients which were admitted to Sahlgrenska University Hospital, Gothenburg, Sweden, in June 2020 during the first wave of infections. The patients presented with severe to critical disease and were hospitalized between 5 and 20 days; samples were collected every 3rd to 7th day during this time (Figure 1A and Table 1) (28, 29). All patients returned for a sampling 3 months after hospitalization and further samples were collected from some patients every third month thereafter. By the 12 months visit in June 2021, the

three patients that remained in the study had all been vaccinated; patient PT1 had received a first dose of Spikevax (Moderna) one week before sampling, patient PT2 had received the first dose of Comirnaty (Pfizer-BioNTech), and patient PT4 had received two doses of Vaxzevria (AstraZeneca). PBMCs were prepared and frozen from blood at time of collection. To maximize information, each sample was split into three: total viable PBMCs, a magnetically depleted fraction enriched for B cells, and sorted antigen specific B cells. Each fraction was labelled with different barcoded hashtag Abs, which allowed for pooling of samples while maintaining the ability to assign each sample and individual (Figure 1A). For enrichment of antigen-specific cells, PBMCs were stained with hashtag Abs and protein-tagged dCODE Klickmers before sorting. Each klickmer featured a unique barcode to allow determination of antigen-specificity and PE to allow flow cytometric sorting. They were loaded with either N, S or RBD antigen, from the infecting Wuhan strain. To minimize the number of unspecific cells, we also included fluorescently labelled N (BV421) and S (APC) tetramers and cells were sorted as Live CD19⁺ Dextramer⁺ and S⁺ or N⁺ (Supplementary Figure 1A). To make sure that only antigen-binding cells would be sorted, initial gates were setup based on IgD⁻ Dextramer⁺ and S⁺ or N⁺ (Supplementary Figure 1A). Pre-pandemic controls showed limited binding to the antigens using this gating strategy (Supplementary Figure 1B). The sorted antigen-specific cells were then pooled with the other two fractions and further stained with TotalSeq Human Universal cocktail, a mixture of 138 barcoded surface Abs. Hence, the transcriptomic, surface antigen phenotype, antigen binding and BCR sequences could be determined from individual samples.

Strict quality filters were applied and the average sequencing depth (represented as UMI per cell) was well over 1000 for all samples, with relatively small differences, indicating good and similar sequence coverage for all samples (Supplementary Figure 2A). Furthermore, the

average number of genes expressed per cell was relatively even and in general over 1000, again consistent to good quality of the scRNA-seq data (Supplementary Figure 2B). Other QC metrics, such as cell complexity and background binding of CITE-Seq Abs, based on seven isotype controls, were evaluated and demonstrated good quality of our data (Supplementary Figure 2C-D). Finally, heatmaps of demultiplexed hashtags and protein bait binding, demonstrated specificity of both (Supplementary Figure 2E-F).

Data was clustered using unsupervised methods based on proteomic and transcriptional data and were visualized in 2D Euclidean space by weighted nearest neighbor (WNN) uniform manifold approximation and projection (UMAP) (Figure 1B and Supplementary Figure 3A-B). The identity of major cell clusters was then determined based on well-known cell-specific markers (Figure 1C-D). To define cell type dynamics we considered only the PBMC pool. Given that not all sampling time points were available for all patients, we grouped them for each patient to allow for proper comparison as follows: Admission (ADM), Hospitalized (3D, 7D, 10D and 14D time points), 3M (months) follow up, Recovered (6M and 9M time points) and Vaccinated (12M time point). The number of total cells varied between samples, due to differences in starting material and recovery, but was still comparable (Supplementary Figure 3C). Time after admission, but not patient sample, impacted the distribution of different cell populations; the frequency of many T cell populations was low at admission, consistent with T cell lymphopenia, but returned to more normal levels already during hospitalization (Figure 1E and Supplementary Figure 3D-E). The proportion of Natural Killer (NK) cells was elevated at admission but decreased to lower levels already after three days. Further, we detected increased proportions of megakaryocytes during hospitalization, consistent with previous reports (7). Finally, among B cells, plasmablasts were elevated during hospitalization and MBC tended to increase after hospitalization, even when excluding antigen-specific, sorted cells

from analysis (Figure 1E). Altogether, the data demonstrated alterations of the circulating immune cell populations during COVID-19 infection that returned to normal levels after recovery.

Enrichment of circulating Th1 CD4 and cytotoxic CD8 T cells early after disease onset

T cells are crucial to limit ongoing intracellular viral infections and to guide B cell responses. To further characterize T cell response, we subsetted the PBMC pool on these populations, re-ran unsupervised clustering analysis and visualized it by WNN-UMAP, using combined protein and RNA data (Figure 2A). The different populations clustered similarly as before. However, recovered T cell number was more variable, as compared to total cells (Supplementary Figure 3F). Indeed, this was marked in admission samples and might be due to early T cell lymphopenia. To gain deeper insights into T cell activation and transcriptional states, we used functional gene sets for T cell subpopulations previously described (30). Based on gene set expression we were able to classify CD8 and CD4 expressing T cells into resting cells, Tregs, interferon (IFN) responding cells and cytotoxic T cells (Figure 2B-C). Interestingly, we identified two clusters of CD4 T cells with a very strong IFN-gene signature. The signature included several IRF genes, IFI genes and other canonical genes associated with IFN response (*MX1*, *STAT1*, *JAK2*). Further analysis revealed this population to be highly expanded at admission, for all patients, and up to 7 days of hospitalization, when it disappeared (Figure 2D-E and Supplementary Figure 3G-H). One other expanded population at admission was cytotoxic CD8 T cells (CTL), characterized by expression of *GZMK*, *GNLY* among others (Figure 2B). Independently of patient, approximately 75% of circulating T cells at admission belonged either to CD8 CTLs or IFN-responding CD4 T cells, consistent with a strong antiviral response (Figure 2F). However, their relative levels returned to normal already 3-7 days after admission and were then maintained at fairly constant levels for at least 1 year after recovery.

Overall, our data suggests a rapid expansion of antiviral T cells in the circulation upon severe COVID-19 infection. CD4 T cells with an IFN gene signature rapidly increase early after infection and return to normal levels within three to seven days during non-lethal infection. In addition, CD8 CTLs are also enriched up to seven days after admission, although in this case there were larger differences between individuals.

Switched MBCs exhibit unique transcriptional features

A key feature of our study is the ability to longitudinally follow antigen specific B cell clones during hospitalization and after recovery. B cells were analyzed separately from other cell types as above and, upon unsupervised re-clustering, twelve B cell clusters were identified (Figure 3A and Supplementary Figure 4A-B). Recovered B cell numbers was comparable among samples (Supplementary Figure 3I). We used gene signature scores to classify these different populations (9) (Figure 3B). Clusters 0, 2, 4, 8, 10 had a strong signature associated with naïve B cells, which included *BACH2*, *ZBTB16*, *APBB2*, *SPRY1*, *TCL1A* and *IKZF2*. Conversely, clusters 1, 3, 5, 6 expressed *CD27*, *CD80*, *CD86*, *TOX*, *TNFRSF13B* and *FCRL5*, core components of the MBC gene signature which was used. Clusters 7 and 11 expressed *CD38*, *IRF4*, *PRDM1*, *XBP1*, confirming their differentiation into antibody secreting plasmablasts. Further, we classified cells by IgM expression or switched immunoglobulin (swIg) signature, which included IgG and IgA expression. Most naïve B cells expressed IgM while the plasmablasts were mostly class switched. Among MBC, we clearly identified clusters 3 and 5 as switched MBC while cluster 1 was mainly composed of IgM-MBC and cluster 6 of IgD/IgM-MBC (Figure 3B-C and Supplementary Figure 4C). The gene expression observations were confirmed when analyzing the isotype from the expressed BCR, using data from the VDJ sequencing (Figure 3C); here we could observe an IgG bias in the swIg memory population and an IgA bias in the plasmablasts (clusters 7 and 11). It should be noted that VDJ sequencing

is not fully able to pick up IgD expressing B cells, due to the primers used, hence the low IgD representation in naïve clusters depicted in Figure 3C.

While the proportion of cells within most of the naïve and MBC clusters were stable over time and between patients, there were significant fluctuations for plasmablast (Figure 3D and Supplementary Figure 3J-K). Plasmablast clusters 7 and 11 were high at admission and during hospitalization, as previously reported (3), but returned to low level after recovery. When considering only the PBMCs and B cell enriched pools, naïve and memory cells were stable throughout the study period (Figure 3D). When only the antigen specific sorted fraction was analyzed, these were clearly enriched in the swlg memory clusters 3 and 5 and increased over time, consistent with constant new memory cell generation (Supplementary Figure 4D).

Mutation frequencies of the different B cell populations, regardless of antigen specificity, indicate that plasmablasts have higher mutation rates than other cell types, despite appearing earlier after hospitalization (Figure 3E). This data is consistent with plasmablasts deriving from rapid differentiation of pre-existing MBC, specific for seasonal CoV, as previously suggested (9-11). In agreement with our gene expression characterization, MBC clusters were mutated while naïve B cell clusters were not. Overall, this analysis demonstrated that antigen-specific switched MBC increase over time, while IgM-MBC remained stable throughout the study period.

To gain better insights into differences between these two MBC subpopulations we compared gene expression between swlg-MBC (clusters 3 and 5) vs IgM-MBC (cluster 1). The differential gene expression analysis revealed significant differences in expressed genes between the two (Supplementary Figure 4E), as previously described using total RNA-seq analysis (31). Genes upregulated in swlg-MBC included integrins *ITGB1* and *ITGB2*, IFN-induced *IFI30*, and genes such as *CD86*, *TNFRSF1B* and several HLA genes, overall suggesting a heightened activation

status. IgM MBC also expressed survival genes, like *MYC/BTG1* and *FOXP1*, and the chemokine receptors *CCR7* and *CXCR4*. Gene ontology (GO) analysis of upregulated genes revealed a much stronger activation profile for swlg-MBC, which included signs of increased signaling, cytokine responses, proliferation, response to IFN, antigen presentation and more (Supplementary Figure 4F). Overall, the data suggests swlg-MBC being in a more activated state, thus possibly having a lower reactivation threshold upon reinfection.

B cells demonstrate differential dynamics depending on antigen specificity

To dissect the evolution of antigen-specific B cell subtypes, we analyzed the antigen-specificity of B cells clusters. This was assessed by looking at unique protein barcodes for protein binding cells, which were indeed highly enriched in the sorted population (Supplementary Figure 4G), as expected. In the following section cells binding to both RBD and S will be defined as S-RBD, while cells binding to S but not RBD will be defined as S. The MBC subset were binding equally to S and to N, with swlg-MBC (clusters 3 and 5) exhibiting a preference for RBD (Figure 4A). While binding was generally lower, antigen binding could be detected among naïve B cells. These could be broadly defined into three groups: clusters 0 and 2 were approximately 50% S and 50% N binding; cluster 4 was mainly composed of S, especially RBD, binders and, finally, all other clusters had a majority of N-binding cells (Figure 4A).

Given the difference in specificity between B cell subpopulations we decided to investigate the dynamics of antigen specific cells over time (Figure 4B, C). We found that among MBC, the proportion of RBD-specific cells continues to increase with time, even in the absence of vaccination. S non-RBD binding MBC had a more complex pattern; their relative frequency decreased during hospitalization but rebound to admission levels by 3 months after infection

and were also slightly boosted upon vaccination (12 month time point). In contrast, N specific MBC were mostly stable throughout the study period.

Longitudinal analysis of antigen specificity and isotypes further highlighted differences between populations (Figure 4D). At admission we could detect a relatively large number of N-specific IgA cells, which correlated with a higher proportion of plasmablasts. At later time points, most patients developed an antigen specific IgG response, concomitant with an increased proportion of antigen-specific MBC, which was slightly more marked for S and RBD specific cells as compared to N. Vaccine administration was particularly efficient in increasing the proportion of isotype switched S- and RBD-specific cells in one patient (PT2), but did not have a strong effect in the other two vaccinated patients.

Together, we could track the longitudinal dynamics of antigen specific B cells over time. We found that switched RBD-binding MBC increased over time, peaking by 9 months. In contrast, S, non-RBD, binding MBC were mostly stable.

Antigen specific B cells possess unique repertoire features

To determine whether antigen-specific cells were indeed enriched for distinct antibody gene usage signatures we first used a Pearson's correlation matrix to analyze Vh gene usage overlap (32). As a control, the analysis was also performed just taking into account the patient and sampling time point and, given the mostly private nature of V gene repertoire (33), samples clustered mainly by patient, regardless of sampling time (Supplementary Figure 5A), as expected. However, when we grouped patients not only by time but also by antigen binding the picture was different (Figure 5A); with the exception of PT4, which maintained a strong private Vh signature, the clustering was mainly dictated by sampling time and antigen specificity. This suggests that antigen-specificity is a strong determinant for Vh selection and that Vh gene usage in antigen-specific B cells was different during hospitalization as compared

to after recovery. This could be explained by the fact that early responses are potentially originating from cross-reactive MBC while later B cells are mainly novel MBC which have matured within the GC.

To better visualize Vh gene usage we generated tile plots where, for each antigen, the top 10 Vh genes were represented proportionally. By grouping by hospitalization status and patient (Figure 5B) we were able to dissect similarities and differences between individuals and proteins. For S non-RBD it has been previously reported that Vh gene usage is skewed towards Vh1-24 (NTD), Vh3-30 (S2) and Vh3-33 in convalescence (34-41). We found that all patients use such genes but in different proportions; for instance, PT2 has strong preference for Vh1-24 and Vh3-33, while PT4 mainly utilizes Vh3-30. Importantly did not find such a strong bias during early response, where only patient PT1 has strong Vh3-30 preference which is lost after recovery. Similarly, Vh3-23 and Vh4-39 have been previously described for N-specific B cells. We did, indeed, find a similar bias in recovered samples, with some patients being more prone to using Vh4-39 while others use Vh3-23. Here the Vh3-23 bias was present in 3 of the 6 patients already at early time point but Vh4-39 usage was detected only in PT4, again arguing for an evolving B cell landscape. For RBD, several Vh genes have been associated with potentially neutralizing mAbs (40, 42, 43) and we detected some of those in our patients (Vh3-23, Vh3-30, Vh1-69, Vh3-53), again with marked difference between samples collected during hospitalization and recovery.

Similarly, patient specific heavy-light chain pairing confirmed that the preference of certain V gene pairings was driven by antigen (Figure 5C and Supplementary Figure 5B). Most patients had Vh3-53 among the top pairings for RBD reactive Abs; however, the light chain pairing was different between patients. PT1 paired Vh3-53 in RBD reactive Abs with Vk3-20, while PT4 paired with Vk1-9. Both these pairings have been already described (40). Recently, a similar

mAb, pairing Vh3-53 with Vk3-20 has been described as a broad neutralizer against viral variants and belonging to the RBD-2 cluster (44). Overall, we detect some similarities between individuals with both time after infection and protein binding being major drivers of Vh gene usage.

Limited clonal expansion and continuous evolution of antigen specific B cells

Next, we wanted to assess when antigen-specific clones arose and for how long they persisted in the memory pool. By sampling longitudinally, we were able to follow the clonal expansion and evolution of antigen specific cells during hospitalization, after recovery and upon vaccination. Mutation rate stratified by antigen binding was steadily increasing over time in all patients, regardless of antigen specificity (Figure 5D and Supplementary Figure 5C). In contrast, the overall mutation rates of B cells were constant throughout the study period (Supplementary Figure 5D). Our observation is consistent with previous data suggesting persistent GC after COVID-19 infection and continued evolution of B cells (12, 19-22). Apart for PT2, we did not detect a significant increase in somatic hypermutation (SHM) after vaccination.

Surprisingly, we observed a very low level of clonal expansion (Figure 5E), even when considering only protein binders (Supplementary Figure 5E). Except for PT3, which had one longitudinal and highly expanded clone, most other patients did not show signs of substantially expanded B cells clones, either early or late. Even during recovery and after vaccination we could not measure significant clonal expansion (Figure 5E). Clonal relationship analysis demonstrated mostly unique clones and few persistent clones across the study period (Supplementary Figure 5F). To better visualize clones that were maintained over time, we focused on clonal families with more than one member (Figure 5F). Clonal analysis on this subset of B cells revealed that most patients maintained at least some clones, which were

generated during infection, in circulation for up to one year (Figure 5F). Importantly, clone related to early plasmablasts and/or early MBC were found to persist for up to one year in the majority of patients (Figure 5F and Supplementary Table 1). However, these were not significantly expanded upon vaccination. These results indicate that most of the stable memory clones are indeed formed at an early stage after infection. When considering antigen-specificity we found that it was easier to assign antigen-binder status at later time points, even for clones within the same family. We reasoned that SHM and affinity maturation might play a role in this phenomenon, with cells at early time points being of too low affinity to bind to the recombinant protein used for sorting.

The data presented here suggest a limited clonal B cell expansion after COVID-19 infection but demonstrate the early formation, persistence and evolution of individual clones for up to one year after infection.

B cells increase antigen binding and neutralizing breadth over time

To further investigate how affinity maturation can influence B cell evolution and specificity, within the same patient, we expressed 30 mAbs derived from all six patients, belonging to 13 clonal families which were found at early and late time points within the same patient (Figure 6A and Supplementary Table 1). These included 5 mAbs which were found to be identical at early and late time points. Two mAbs (9 and 16) failed to be expressed.

First, we tested all available sera (Figure 6B) and mAbs (Figure 6C) for binding to a panel of viral antigens, using a multiplex electrochemiluminescence-based assay. The first panel included full length S, as well as RBD and N terminal domains (NTD) of S and N from the SARS-CoV-2 Wuhan strain; S from SARS-CoV-1, MERS, HKU-1 and OC43, as well as H3 from Influenza A virus/Hong Kong. The second panel included S proteins from several SARS-CoV-2 variant strains: Wuhan, Alpha (B.1.1.7), Beta (B.1.351), Gamma (P.1), Delta (B.1.617.2, AY.4 and

AY.4.2) and Omicron (B.1.1.529 BA.1). Importantly, all our patients were hospitalized in June 2020 and thus infected with the ancestral strain. The serum Ab binding profile was similar for all patients and showed fairly strong binding to most of the antigens. Sera collected at admission was also reactive to seasonal CoV for five of six patients (Figure 6B). SARS-CoV-2 specific IgG developed within few days, reached a peak by three months in most patients and levels were then relatively stable throughout the study period. We could detect a sharp increase in IgG Abs binding to all S but not N or control influenza antigens after vaccination (Figure 6A, 12 month time point). By comparing this with the B cell data we can speculate that after vaccination many MBC directly differentiate into Ab secreting plasma cells.

The majority of the expressed mAbs was able to bind at least one of the tested proteins, confirming the specificity of our sorting approach (Figure 6C). In general, binding to N was harder to assign as several strong S-specific mAbs also showed N signal, including the CR3022 Ab used as binding reference (45). Therefore, we assigned N specificity only to mAbs with stronger N signal, as compared to other proteins. Four mAbs were specific for the N protein (mAbs 3, 4, 5 and 17). mAb 18, which was identical at early and late time points, was specific for NTD of S. This specific BCR, of the IgA1 isotype, was present at the 7 days, 10 days and 3 months sampling time points. This mAb did not cross react with seasonal CoV but had some minor cross-reactivity with MERS-S and Alpha-S. For family PT4_2382 (mAbs 10 to 15) we could not confidently assign a protein specificity. For S specific cells, the large majority of cloned mAbs were RBD directed. Importantly, we could detect that cells, clonally related to the early plasmablasts, re-emerged at later time points as MBC. This was the case for one of our clonal families (with mAbs 22 to 26), belonging to patient PT5. Here mAbs 22 and 23 were identified at admission from plasmablasts, showing strong RBD binding and an already high mutation rate of 7-8%. At the 3 and 9 months follow-up, clonal relatives were then identified

in the MBC pool with an increased mutation rate of 7-11%. Importantly, mAbs from the memory pool (mAbs 24, 25, 26) showed significantly higher antigen binding breadth, including some binding to Omicron BA.1. All these mAbs were originally IgA1. The data here suggests that clonal relatives of early plasmablasts also entered the GC and there continued to evolve as MBC with further binding breadth, or that the same GC clones can generate early plasmablasts and late MBC. When tested on the S variant panel we analyzed the response for every clonal family. Apart from the pair mAb20 and 21, most S-specific families consistently increased breadth over time, regardless of vaccination status. While most families, eventually acquired binding to all variants, including Beta and Delta, binding to Omicron BA.1 was not obtained by all and was, generally, of lower magnitude. Specificity of the highly sensitive electrochemiluminescence approach was confirmed using traditional ELISAs. As expected, some interactions fell below detection in this case, giving less background, yet the ELISA in general corroborated our the results and demonstrated a level of polyreactivity for some of the mAbs (Supplementary Figure 6A-B).

To verify whether the increase of binding breadth correlated with increased neutralizing activity against viral variants, we tested the mAbs using HEK293T-ACE2 cells and SARS-CoV-2 pseudotyped lentivirus (46). Of the S-binding families, mAbs mAb01 and mAb02 and mAb27 and mAb28 did not neutralize, despite an increase in binding breadth (Figure 6C and Supplementary Figure 6C). mAb06 was poorly neutralizing but its later relative mAb07 was potent against ancestral SARS-CoV2 S as well as against Alpha, Beta, Gamma and Delta (Supplementary Figure 6C-E). The family from which we had most binding clones, PT5_323, showed increasing neutralization breadth. Of the mAbs isolated at admission, mAb22 did not neutralize, while mAb23 was poorly neutralizing against ancestral S. The 3 months relative, mAb24, had higher IC_{50} against ancestral S, similar neutralization capacity against Alpha and

some weak neutralization against Beta, Gamma and Delta. The 9 months mAbs, mAb25 and mAb26, were potent against all strains, except Omicron BA.1 (Supplementary Figure 6C-E).

In summary, we demonstrated a continued evolution of B cells, irrespective of vaccination, with B cells acquiring increased antigen binding and neutralization breadth over time.

DISCUSSION

Despite a very successful vaccination campaign, SARS-CoV-2 is still circulating and causing significant disease (1). Viral variants have continued to emerge (47-50), and there is need to better understand the development of B cell immunity after infection and vaccination. While levels of neutralizing Abs in sera are a good correlate of protection from infection, circulating MBC, with varying specificity and breadth, can also be rapidly reactivated by reinfection (51, 52). Even if the specific B cells are not able to protect from infection, MBC can rapidly differentiate to plasmablasts or re-enter germinal centers, thus fighting the virus, limiting viral replication and decreasing disease severity. Many studies have followed serum Ab activity in patients longitudinally, while less is known about the dynamics and specificity of MBC during disease and convalescence (53-56). A few studies have analyzed the clonal relationship of B cells, but these usually start with samples collected one month after disease onset, at the earliest, excluding any comparisons to acute plasmablast responses (12, 19-22). A study by Dugan *et al.* used a technical setup similar to ours; however, the majority of the patients were not followed longitudinally but rather sampled cross-sectionally, either during hospitalization or after recovery (9). To the best of our knowledge, a detailed study where the transcriptional dynamics of total PBMCs and B cells followed longitudinally during disease, convalescence and after vaccination in individual patients has not previously been performed. This rather unique setting allowed us to track responses in few but densely sampled individuals to

elucidate the dynamics of antigen specific B cells, as well as more general immune responses, after COVID-19 infection.

In our study we analyzed samples from six patients with severe/critical COVID-19 infection and we report significant alterations in the circulating immune populations during disease. As described previously (5), we demonstrated the presence of CD4 T cells with strong IFN signature early during hospitalization which disappeared already by 3 months. Other significant alterations in the immune cells composition or transcriptional features normalized by day 14 after admission. Given the low number of patients it is not possible to make any solid conclusion regarding the relationship between the analyzed immune responses and comorbidities and general immune status. Three of our patients (PT2, PT3 and PT4) had respiratory comorbidities, however, they did not have any immediately noticeable difference in the frequency of the different immune cell populations or B cells. PT2 had partial IgA deficiency (due to anti-IgA IgGs) and decreased IgG2 serum levels; however, the frequency of MBC and antigen-specific MBC were normal and even higher than others. In general, we can conclude that, regardless of comorbidities, immune-status, age and sex, immune responses showed similar features in all patients.

The main goal of the study was to investigate longitudinal B cell clonal dynamics and antigen specificity. Here, we reveal that upon infection, patients do not present largely expanded B cell clones in the blood. Rather, several smaller clones can be detected at all sampled time points, even among early plasmablasts, and these do persist in circulation for at least one year, even after vaccination. Further, mAbs within the same patient, belonging to the same clonal family, acquired increased binding to several viral variants over time, revealing persistent ongoing clonal affinity maturation.

A significant observation is the expansion of MBC after recovery, with an increased frequency of antigen specific B cells and higher somatic hypermutation across binders. However, this was not significantly accelerated by vaccination. Indeed, the SHM rate, in antigen specific cells, was similar between 9 months after infection and 12 months (which was after administration of one or two vaccine doses). This suggests that, in these patients, vaccination was not sufficient to elicit a strong clonal expansion and reactivation of infection-derived MBC, at least when sampled from the circulating PBMCs. In contrast, data from sera indicates a strong differentiation of MBC to Ab secreting plasmablasts in response to vaccination. We could not detect antigen binding for the majority of plasmablasts due to their low surface Ig expression, but most of those we could identify were RBD specific. This, combined with the already high mutation rate and the predominant IgA and IgG isotype expression confirms that these most likely originate from seasonal CoV specific MBC which were rapidly reactivated (9-11). While many of these plasmablasts did not have clonal relatives at later time-points, there were notable exceptions, demonstrating that after activation, clonal relatives to these early plasmablasts must have entered GC and acquired further mutations and breadth before differentiating to MBC. An alternative, possible explanation is that early plasmablasts were in fact generated from naïve B cells forming COVID-induced GC, although this is not fully compatible with the high mutation rates of these cells.

When comparing IgM and switched MBC for their transcriptional signatures we detected significant differences. Some of the genes identified in our study are similar to previously identified sequences using bulk RNAseq (31). Further, our GO analysis highlighted that swIg-MBC show a more “activated” transcriptional signature as compared to IgM MBC, despite having similar mutation rates. Further, swIg-MBC preferentially bound RBD while memory IgM had more N-binding. This might imply that RBD-specific swIg-MBC are primed to get

reactivated upon reinfection. Increased RBD binding in the MBC population was previously reported (12) but we add to the data by showing that vaccination does not dramatically increase this targeting.

We were surprised by the limited clonal expansion after disease and vaccination. While previous studies (12, 20) reported similar findings, their first sampling point was approximately a month after infection. A more comprehensive BCR repertoire sequencing also indicated lack of clonal expansion after mild disease (26) but high clonal expansion starting a month after vaccination. Our data shows that individuals in our study had a more diverse response after vaccination, given their pre-existing Ab immunity, as recently suggested in another study (57). Despite this, we could identify a number of antigen-specific and persistent clones together with the emergence of novel, antigen-binding B cells at later time points. Importantly, like essentially all single cell studies, our work has very limited sampling depth of the B cell receptor repertoire (58). This limits some of the conclusion we can make from our analyses, but make it even more remarkable that we were still able find persisting clones even within this limited repertoire space. It is possible that more such clones would be revealed had the sequencing depth been greater. While clonal expansion is undoubtedly vigorous in animal models upon vaccination or viral infection, the situation in humans is clearly more complex. Animal studies suggest that GC have the ability to maintain a diverse response (59, 60), but how this is attained is not completely clear. With B cell activation after COVID-19 infection initially consisting of recalled seasonal CoV-specific B cells, it could be that such B cells, in combination with pre-existing serum Abs, could promote the diversity of the ongoing GC. A more diverse GC would favor novel responses and the ability of the B cells to adapt to existing and future viral variants (15, 17, 61).

The main limitation of this study is the relatively low number of analyzed patients. This makes it difficult to link some of the gene expression and cellular data to the clinical status of the patients. However, given the type of analysis performed on the patients and our wish to follow immune responses over time, we preferred to have a smaller cohort of regularly sampled patients rather than larger number of individuals with less frequent sampling. Stronger responses may indeed have appeared after a second or third dose of vaccine. Given the variability between individuals, future studies including more patients, especially after vaccination, would be needed to confirm our results. Further, having started our study before emergence of the occurrence of dominant viral variants, we only included Wuhan-Spike as our sorting antigen. Inclusion of S proteins from viral variants in the sorting and sequencing procedure would have been a valuable addition to better understand future protection in the individuals. Finally, while we demonstrate binding breadth and neutralization capacity of the cloned mAbs, it is impossible to know if this represents their *in vivo* protective capacity. Future studies will also need to address the presence and persistence of antigen-specific B cells in tissues and their clonal relationship with circulating B cells. A previous study reported persistence of immune cells in organs, up to 6 months after COVID19 infection (62) but their clonality with circulating B cells was not investigated. Fine needle aspirates of LN (63, 64) could give important insights into this issue but the optimal, although technically impossible, sampling site would be the lungs, where tissue resident BMDM might be the first line of protection against viral reinfection.

Overall, we show a number of persisting, SARS-CoV-2 specific clones which are first elicited early after infection and maintained up to one year, while increasing their binding and neutralization breadth. Given the emergence of viral variants and successful deployment of

vaccines, future studies should address how and if similar clones can be reactivated upon vaccination or reinfection.

METHODS

Sampling

The study cohort was recruited from hospitalized COVID-positive patients at the Sahlgrenska University Hospital, Gothenburg, Sweden. Pre-pandemic controls were obtained from healthy donors seronegative by ELISA to S and N.

Whole blood samples and serum samples of patients were drawn at admission and every three days during hospitalization. After discharge, samples were drawn every 3 months for up to one year. Whole blood samples were collected in lithium heparin tubes, from which PBMCs were isolated with Lymphoprep (Stemcell Technologies) and stored in liquid nitrogen. Serum samples were collected in serum tubes and stored at -80°C (28, 29).

Preparation of immune cell populations

For scRNAseq, PBMCs were thawed and divided in three fractions, labelled with barcoded Hashtag Antibodies 1 – 6 (Biolegend, catalogue nr. 394661, 394663, 394665, 394667, 394669, 394671), in order to facilitate later demultiplexing, and submitted to different downstream handling. The starting number of cells was approximately 10×10^6 cells/tube. For the first group, 300 000 total PBMCs only underwent dead cell depletion (EasySep Dead Cell Removal (Annexin V) Kit - Stemcell Technologies). In the second group, B cells were isolated from 700 000 cells of the crude sample via negative selection (EasySep Human Pan-B Cell Enrichment Kit - Stemcell Technologies), prior to dead cell removal. Both fractions were then incubated with Human TruStain FcX (Biolegend, catalogue nr., 422302) before being stained with dCODE Dextramer-PE (Immunodex, product code dCXC) complexes containing SARS-CoV-2 Nucleoprotein, Spike or the Receptor Binding Domain (RBD) of SARS-CoV-2 Spike (Sino Biological, catalogue nr. 40588-V27B-B, 40591-V27H-B, 40592-V27H-B). The remaining PBMC

fraction was used for isolation of SARS-CoV-2-specific B cells using fluorescence-activated Cell Sorting (FACS)(see below).

Fluorescence-activated Cell Sorting

To test the specificity of our approach, COVID-positive and pre-pandemic controls were stained with dCODE Dextramer-PE complexes, as well as complexes of SARS-CoV-2 Spike and Nucleoprotein coupled with fluorophore-streptavidin conjugates (APC, Biolegend, catalogue nr. 405207). Following addition of anti-human CD19 APC-H7 (BD Biosciences, catalogue nr. 115530), IgD BUV395 (BD Biosciences, catalogue nr. 563813) and LIVE/DEAD Fixable Aqua Dead Cell Stain (ThermoFisher, catalogue nr. L34966) cells were acquired on BDFACSAria Fusion X20.

For sorting of SARS-CoV-2-specific B cells, PBMCs were stained with TotalSeq anti-human Hashtag Antibody and Human TruStain FcX (Biolegend), similarly to the other two fractions. Cells were subsequently stained with dCODE Dextramer-PE complexes, as well as complexes of SARS-CoV-2 Spike or Nucleoprotein coupled with fluorophore-streptavidin conjugates (APC and BV421, Biolegend, catalogue nr. 405225) in order to reduce background binding. Following addition of anti-human CD19 APC-H7, IgD BUV395 and LIVE/DEAD Fixable Aqua Dead Cell Stain, cells were sorted using a BDFACSAria Fusion. The whole sample was sorted.

Single cell RNA sequencing

After being processed, the three immune cell populations fractions were stained with TotalSeq-C Human Universal Cocktail, V1.0 (Biolegend, catalogue nr. 399905), as well as TotalSeq anti-human CD72, IgG Fc, CD197 (CCR7), CD45RB, CD193 (CCR3), TCR γ/δ , LAIR1 (CD395) and CD366 (Tim-3) (Biolegend, catalogue nr. 316211, 410727, 353251, 310211, 310733, 331231, 342807 and 345049). Cells were pooled, washed using the Laminar Wash Mini

system (Curiox Biosystems) before using the Chromium Next GEM Single Cell 5' Kit v2 and Chromium Next GEM Chip K Single Cell Kit (10x Genomics).

Libraries were created using the Library Construction Kit, 5' Feature Barcode Kit, Chromium Single Cell Human BCR Amplification Kit, Dual Index Kit TT Set A (PN-1000215) and Dual Index Kit TN Set A (PN-1000250) (10x Genomics). Their quality and quantity was assessed using the Agilent Tapestation system and the Qubit Fluorometer (Invitrogen), before being sent for sequencing on Illumina NovaSeq 6000 as per the instructions provided in 10X Genomics user guide for the read length and depth. Sequencing was performed by SNP&SEQ Technology Platform, Science for Life Laboratory (Uppsala Biomedical Centre, Uppsala University, Sweden).

Data analysis

Raw fastq files were processed through the 10x cellranger pipeline using the multi command and default parameters with reference genome GRCh38-2020-A. Raw UMI count matrices generated from the cellranger 10X pipeline were loaded and merged into a single Seurat object. Cells were discarded if they met any one of the following criteria: percentage of mitochondrial counts > 15%; number of unique features or total counts was in the bottom or top 0.5% of all cells; number of unique features < 50. Red cells were discarded by filtering on *HBB*, *HBD*, *HBA1*, *HBA2*, *HBM*, *HBQ1* expression.

Gene counts were normalized to the same total counts per cell (1000) and natural log transformed (after the addition of a pseudocount of 1). The normalized counts in each cell were mean-centered and scaled by their standard deviation, and the following variables were regressed out: number of features, percentage of mitochondrial counts, and the difference between the G2M and S phase scores. Surface protein and antigen probe were normalized by a centered log-ratio (CLR) normalization.

Data integration across cells originating from different samples, was done using the Anchor method within Seurat. To integrate surface protein data and transcriptomic data the FindMultiModalNeighbors function in Seurat was used to construct a weighted nearest neighbor (WNN) graph. For each cell, the nearest neighbors based on a weighted combination of two modalities was identified and used as input for dimensional reductions. Selection of the number of components for the nearest-neighbor network computation was based on their visualization in an elbow plot. Uniform manifold approximation and projection (UMAP) (65) was performed for the spatial visualization of the single- cell dataset and features and cell clustered using the Louvain algorithm.

For CITE-Seq we took advantage of 7 internal isotype controls included in the mixture (MouseIgG1, MouseIgG2a, MouseIgG2b, RatIgG2b, RatIgG1, RatIgG2a and HamsterIgG) and performed QC using the scater and DropletUtils packages (66, 67). Briefly, the isotype controls were used as measure of non-specific protein aggregates and as a measure to assess the specificity of our protein labels.

Protein binding reads were normalized using the CLR function for each sample. To define probe binding cells we use the HTODemux function in Seurat. All positive cells were subsequently examined and cells positive only for only one probe were assigned the corresponding specificity.

Cells which did not have an associated protein barcode or which had multiple barcodes were defined as non-binding. RBD binding cells were defined as cells barcoded with both S and RBD. Interestingly there was a very clear population of cells just having RBD barcodes assigned; given that we could not confidently assign those cells a clear specificity we defined them as non-binders. Therefore, cells defined as S specific, included only non-RBD epitopes, while RBD-cells were S binders with RBD specificity.

BCR sequences data processing

The BCR sequence data was processed using the Immcantation toolbox (v4.0.0) using the IgBLAST and IMGT germline sequence databases, with default parameter values unless otherwise noted. The IgBLAST database was used to assign V(D)J gene annotations to the BCR FASTA files for each sample using the Change-O package (68), resulting in a matrix containing sequence alignment information for each sample for both light and heavy chain sequences.

BCR sequence database files associated with the same individual were combined and processed to infer the genotype using the TIgGER package (69) as well as to correct allele calls based on the inferred genotype. The SHazaM package (68) was used to evaluate sequence similarities based on their Hamming distance and estimate the distance threshold separating clonally related from unrelated sequences. For each patient, Ig sequences were assigned to clones using the hierarchicalClones function in Scoper (70), where the distance threshold was set to the corresponding value predicted with SHazaM in the previous step. Clonal assignment was done based on both heavy and light chain and cells with multiple heavy or light chains were excluded. Germline sequences were generated for each patient using the genotyped sequences (FASTA files) obtained using TIgGER (69). BCR mutation frequencies were then estimated using SHazaM. The BCR sequence data, clone assignments, and estimated mutation frequencies were integrated with the single-cell RNA-seq data by aligning and merging the data with the metadata slot in the processed RNA-seq Seurat object.

Differential gene expression analysis

Differentially expressed genes between different clusters, were identified using the FindAllMarkers function from Seurat using default settings (Wilcoxon test and Bonferroni p value correction). Significant genes with average log fold change > 0.25 and expressed in $>25\%$

of cells in that group were ranked according to fold change. For comparison between two groups the FindMarkers function was used and all genes represented on a VolcanoPlot.

Gene Ontology (GO) analysis

Differentially expressed genes between two populations were identified as above and analyzed using the ClueGo plugin in Cytoscape (71, 72). Upregulated genes for each population were analyzed using the GOBiological and GOImmune terms with minimum 4 genes and 4% settings. Only pathways with p value < 0.001 were considered. Common pathways were consolidated by clueGO. Pathways were exported and plotted.

Selection and Expression of mAbs

Clonal families present at multiple time points were filtered and sequences manually inspected. Abs were selected based on binding characteristics, cell type, mutation rate and isotype. Both heavy and light chains were synthesized and mAbs expressed as IgG1 in small scale by BioIntron. After testing, further production was carried out in-house as follows. To express recombinant antibodies plasmids encoding corresponding heavy and light chains were mixed in equal ratio. Transfection of Expi293 cells was carried out by ExpiFectamine 293 Transfection Kit (ThermoFisher) according to the manufacturer instruction. After four to seven days supernatants were collected and filtered. Purification of the immunoglobulins was carried out by Akta Start System (GE Healthcare) using protein G column. Elution of bound antibodies was done by 0.1 M glycine buffer, pH 2.7. To neutralize the solution coming from the column collecting tubes contained 1 M Tris buffer, pH 9.0. Antibody-containing eluates were concentrated by centrifugation through VivaSpin columns with a 30 kDa cut-off. Estimation of antibody concentration was done by NanoDrop (ThermoFisher) equipment.

Electrochemiluminescence assay

Serum samples were analyzed for IgG, IgA and IgM Abs binding to SARS-CoV-2 proteins. Additionally, sequenced BCR repertoires of the patients were used to design monoclonal human IgG1 antibodies (mAbs) (produced by Biointron).

Binding to a panel of respiratory viruses was analyzed using a multiplex electrochemiluminescence assay including SARS-CoV-2 S, RBD, NTD and N antigens, as well as SARS-CoV-1, MERS, OC43 and HKU1 S antigens and Influenza H3 antigen (V-PLEX COVID-19 Coronavirus Panel 1 (IgG, IgM, IgA) Kit, Meso Scale Discovery, Rockville, USA). SARS-CoV-2-specific antibodies were further analyzed for binding to S antigen of several SARS-CoV-2 variants (V-PLEX SARS-CoV-2 Panel 23 (IgG) Kit, Meso Scale Discovery, Rockville, USA). For read-out, we used the Meso Quickplex SQ 120 reader (Meso Scale Discovery, Rockville, USA). Serum and plasma samples were analyzed at a 1:5,000 and/or 1:50,000 dilution and mAbs were analyzed at 2-4 concentrations ranging from 0.01 ug/ml to 10 ug/ml. For analysis, binding signal was log transformed and normalized, using the normalize function in Prism (GraphPad), with the top binder set to 1.

ELISA

Microton medium binding half-well ELISA plates (Greiner Biotech) were coated overnight at 4°C with recombinant SARS-CoV-2 N (1µg/ml), SARS-CoV-2 Spike, SARS-CoV-2 Spike RBD, or IAV HA H3 Brisbane (2µg/ml), in 25ul PBS. Plates were blocked with 25ul PBS/2% BSA for 1 hours at RT. After 3x washes with PBS+0.05% Tween-20 (PBST) using a BioTek 405 LS Plate Washer, plates were incubated with three-fold dilutions of mAbs starting from 30µg/ml in PBST for 1h at RT. After 3x washes plates were incubated with 25ul of anti-human IgG (Southern Biotech, catalogue nr. 2048-05) diluted 1:6000 for 1h at RT. After 3x washes, plates

were developed for 5 minutes using 1-step Ultra TMB (Thermo Fisher) and halted with 2M H₂SO₄. Plates were read at 450nm with a TECAN Plate reader.

Pseudotyped virus neutralization assays

Pseudotyped virus neutralization assays were performed as previously described (46). Briefly, pseudotyped lentiviruses displaying spikes (with C-terminal truncations) from the ancestral variant or from variants of concern and packaging a firefly luciferase reporter gene were generated by co-transfecting HEK293T cells using Lipofectamine 3000 (Invitrogen) according to the manufacturer's protocols. Pseudotyped virions standardized to an input producing ~100,000 RLU were incubated with serial dilutions of recombinant antibodies for 60 min at 37°C prior to the addition of ~15,000 HEK293T-ACE2 cells and incubation for 48h and luminescence was measured using Bright-Glo (Promega) on a GM-2000 luminometer (Promega).

Statistical Analysis

For analysis of single cell data relevant statistical analyses are indicated in the respective method section. For comparisons of immune cell proportion ordinary one-way ANOVA with post-hoc Tukey correction for multiple comparison was used. Outliers were identified using Grubbs' method. All statistical comparisons were performed using the rstatix package in R.

Data Availability

The processed scRNA-seq data reported in this paper are available in the ArrayExpress database (<http://www.ebi.ac.uk/arrayexpress>) under accession number: E-MTAB-12392. All codes supporting the current study are available from the corresponding author upon request.

Study Approval

The study was approved by the Swedish Ethics Review Authority (Registration number 2020–01771). Written informed consent was received from all patients, prior to participation

AUTHOR CONTRIBUTIONS

Conceptualization, M.B., D.A.; methodology, L.S., H.A., A.E., N.R.M., D.J.S., I.V.S., B.M., A.L., M.B., D.A.; formal analysis, L.S., H.A., A.E., D.J.S., A.L., M.B., D.A.; investigation, L.S., H.A., A.E., N.R.M., P.I., I.V.S., D.J.S.; data curation, D.A.; resources, S.L., I.V.S., N.M., L-M.A., M.G., B.M.; patient inclusion and sampling, L.M.A, M.G.; project administration, A.L., M.B., D.A.; supervision, B.M., A.L., M.B., D.A.; funding acquisition, A.L., M.B., D.A., M.G.; writing – original draft, D.A.; writing – review & editing, L.S., H.A., A.E., N.R.M., D.J.S., S.L., P.I., I.V.S., N.M., L-M.A., M.G., B.M, A.L., M.B., D.A.

ACKNOWLEDGMENTS

We would like to thank the patients, doctors, nurses and additional staff at Sahlgrenska University Hospital involved in the study. We would like to thank Nicholas Borcharding (Washington University St. Louis, USA) for help with some codes used to generate figures and Victor Greiff (University of Oslo, Norway) for helpful discussion on BCR analysis. The project was supported by grants from the European Research Council (ERC-StG, B-DOMINANCE, grant nr. 850638 to D.A.), the Swedish Research Council (grant nr. 2017-01439 to D.A. and 2021-06545 to M.G.), grants from the SciLifeLab National COVID-19 Research Program, financed by the Knut and Alice Wallenberg Foundation (KAW2020.0182 to D.A., M.B., and A.L. and & 2020.0241 to M.G.), by the Swedish state, under an agreement between the Swedish government and the county councils (ALF agreement ALFGBG-965885 to M.G.) and Sahlgrenska Academy (GU2020/2376-JNR2020 to D.A.). N.R.M. is supported by a postdoctoral grant from the Swedish Society for Medical Research (SSMF, grant nr. PD20-0017). Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is

part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory. The SNP&SEQ Platform is also supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation. Graphical abstract was designed using BioRender.

REFERENCES

1. Dong E, et al. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis.* 2020;20(5):533-4.
2. Collaborators C-EM. Estimating excess mortality due to the COVID-19 pandemic: a systematic analysis of COVID-19-related mortality, 2020-21. *Lancet.* 2022;399(10334):1513-36.
3. Arunachalam PS, et al. Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science.* 2020;369(6508):1210-20.
4. Koutsakos M, et al. Integrated immune dynamics define correlates of COVID-19 severity and antibody responses. *Cell Rep Med.* 2021;2(3):100208.
5. Mathew D, et al. Deep immune profiling of COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science.* 2020;369(6508).
6. Schulte-Schrepping J, et al. Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell.* 2020;182(6):1419-40 e23.
7. Bernardes JP, et al. Longitudinal Multi-omics Analyses Identify Responses of Megakaryocytes, Erythroid Cells, and Plasmablasts as Hallmarks of Severe COVID-19. *Immunity.* 2020;53(6):1296-314 e9.
8. Sterlin D, et al. IgA dominates the early neutralizing antibody response to SARS-CoV-2. *Sci Transl Med.* 2021;13(577).
9. Dugan HL, et al. Profiling B cell immunodominance after SARS-CoV-2 infection reveals antibody evolution to non-neutralizing viral targets. *Immunity.* 2021;54(6):1290-303 e7.
10. Ng KW, et al. Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science.* 2020;370(6522):1339-43.
11. Anderson EM, et al. Seasonal human coronavirus antibodies are boosted upon SARS-CoV-2 infection but not associated with protection. *Cell.* 2021;184(7):1858-64 e10.
12. Sokal A, et al. Maturation and persistence of the anti-SARS-CoV-2 memory B cell response. *Cell.* 2021;184(5):1201-13.e14.
13. Song G, et al. Cross-reactive serum and memory B-cell responses to spike protein in SARS-CoV-2 and endemic coronavirus infection. *Nature communications.* 2021;12(1):2938.
14. Kaplonek P, et al. Early cross-coronavirus reactive signatures of humoral immunity against COVID-19. *Sci Immunol.* 2021;6(64):eabj2901.
15. Mathew NR, et al. Single-cell BCR and transcriptome analysis after influenza infection reveals spatiotemporal dynamics of antigen-specific B cells. *Cell reports.* 2021;35(12):109286.
16. Wong R, et al. Affinity-Restricted Memory B Cells Dominate Recall Responses to Heterologous Flaviviruses. *Immunity.* 2020;53(5):1078-94.e7.
17. Baumgarth N. How specific is too specific? B-cell responses to viral infections reveal the importance of breadth over depth. *Immunological reviews.* 2013;255(1):82-94.
18. Purtha WE, et al. Memory B cells, but not long-lived plasma cells, possess antigen specificities for viral escape mutants. *J Exp Med.* 2011;208(13):2599-606.
19. Gaebler C, et al. Evolution of antibody immunity to SARS-CoV-2. *Nature.* 2021;591(7851):639-44.
20. Sakharkar M, et al. Prolonged evolution of the human B cell response to SARS-CoV-2 infection. *Sci Immunol.* 2021;6(56).

21. Wang Z, et al. Naturally enhanced neutralizing breadth against SARS-CoV-2 one year after infection. *Nature*. 2021;595(7867):426-31.
22. Muecksch F, et al. Affinity maturation of SARS-CoV-2 neutralizing antibodies confers potency, breadth, and resilience to viral escape mutations. *Immunity*. 2021;54(8):1853-68 e7.
23. Goel RR, et al. mRNA vaccines induce durable immune memory to SARS-CoV-2 and variants of concern. *Science*. 2021;374(6572):abm0829.
24. Rothausler K, and Baumgarth N. B-cell fate decisions following influenza virus infection. *European journal of immunology*. 2010;40(2):366-77.
25. Yewdell WT, et al. Temporal dynamics of persistent germinal centers and memory B cell differentiation following respiratory virus infection. *Cell Rep*. 2021;37(6):109961.
26. Kotagiri P, et al. B cell receptor repertoire kinetics after SARS-CoV-2 infection and vaccination. *Cell Rep*. 2022;38(7):110393.
27. Zhang Y, et al. Analysis of B Cell Receptor Repertoires Reveals Key Signatures of the Systemic B Cell Response after SARS-CoV-2 Infection. *J Virol*. 2022;96(4):e0160021.
28. Marklund E, et al. Serum-IgG responses to SARS-CoV-2 after mild and severe COVID-19 infection and analysis of IgG non-responders. *PloS one*. 2020;15(10):e0241104.
29. Patel H, et al. Proteomic blood profiling in mild, severe and critical COVID-19 patients. *Scientific reports*. 2021;11(1):6357.
30. Szabo PA, et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature communications*. 2019;10(1):4706.
31. Seifert M, et al. Functional capacities of human IgM memory B cells in early inflammatory responses and secondary germinal center reactions. *Proc Natl Acad Sci U S A*. 2015;112(6):E546-55.
32. Greiff V, et al. Systems Analysis Reveals High Genetic and Antigen-Driven Predetermination of Antibody Repertoires throughout B Cell Development. *Cell Rep*. 2017;19(7):1467-78.
33. Galson JD, et al. In-Depth Assessment of Within-Individual and Inter-Individual Variation in the B Cell Receptor Repertoire. *Front Immunol*. 2015;6:531.
34. Wang Z, et al. Analysis of memory B cells identifies conserved neutralizing epitopes on the N-terminal domain of variant SARS-Cov-2 spike proteins. *Immunity*. 2022;55(6):998-1012 e8.
35. Voss WN, et al. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science*. 2021;372(6546):1108-12.
36. Suryadevara N, et al. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell*. 2021;184(9):2316-31 e15.
37. McCallum M, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell*. 2021;184(9):2332-47 e16.
38. Li D, et al. In vitro and in vivo functions of SARS-CoV-2 infection-enhancing and neutralizing antibodies. *Cell*. 2021;184(16):4203-19 e32.
39. Cerutti G, et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell host & microbe*. 2021;29(5):819-33 e7.
40. Chen EC, et al. Convergent antibody responses to the SARS-CoV-2 spike protein in convalescent and vaccinated individuals. *Cell reports*. 2021;36(8):109604.

41. Wang Y, et al. A large-scale systematic survey reveals recurring molecular features of public antibody responses to SARS-CoV-2. *Immunity*. 2022;55(6):1105-17 e4.
42. Robbiani DF, et al. Convergent antibody responses to SARS-CoV-2 in convalescent individuals. *Nature*. 2020;584(7821):437-42.
43. Yuan M, et al. Structural basis of a shared antibody response to SARS-CoV-2. *Science*. 2020;369(6507):1119-23.
44. Windsor IW, et al. Antibodies induced by ancestral SARS-CoV-2 strain that cross-neutralize variants from Alpha to Omicron BA.1. *Sci Immunol*. 2022:eabo3425.
45. ter Meulen J, et al. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. *PLoS Med*. 2006;3(7):e237.
46. Sheward DJ, et al. Beta RBD boost broadens antibody-mediated protection against SARS-CoV-2 variants in animal models. *Cell Rep Med*. 2021;2(11):100450.
47. Harvey WT, et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nature reviews Microbiology*. 2021;19(7):409-24.
48. Faria NR, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021;372(6544):815-21.
49. Tegally H, et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv*. 2020:2020.12.21.20248640.
50. Tegally H, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021;592(7854):438-43.
51. Bhattacharya D. Instructing durable humoral immunity for COVID-19 and other vaccineable diseases. *Immunity*. 2022;55(6):945-64.
52. Sette A, and Crotty S. Immunological memory to SARS-CoV-2 infection and COVID-19 vaccines. *Immunological reviews*. 2022.
53. Dan JM, et al. Immunological memory to SARS-CoV-2 assessed for up to 8 months after infection. *Science*. 2021;371(6529).
54. Cohen KW, et al. Longitudinal analysis shows durable and broad immune memory after SARS-CoV-2 infection with persisting antibody responses and memory B and T cells. *Cell Rep Med*. 2021;2(7):100354.
55. Yang Y, et al. Longitudinal analysis of antibody dynamics in COVID-19 convalescents reveals neutralizing responses up to 16 months after infection. *Nat Microbiol*. 2022;7(3):423-33.
56. Rodda LB, et al. Functional SARS-CoV-2-Specific Immune Memory Persists after Mild COVID-19. *Cell*. 2021;184(1):169-83 e17.
57. Schaefer-Babajew D, et al. Antibody feedback regulation of memory B cell development in SARS-CoV-2 mRNA vaccination. *medRxiv*. 2022:2022.08.05.22278483.
58. Greiff V, et al. Bioinformatic and Statistical Analysis of Adaptive Immune Repertoires. *Trends Immunol*. 2015;36(11):738-49.
59. Kuraoka M, et al. Complex Antigens Drive Permissive Clonal Selection in Germinal Centers. *Immunity*. 2016;44(3):542-52.
60. Tas JM, et al. Visualizing antibody affinity maturation in germinal centers. *Science*. 2016;351(6277):1048-54.
61. Baumgarth N. The shaping of a B cell pool maximally responsive to infections. *Annual Review of Immunology*. 2021.

62. Poon MML, et al. SARS-CoV-2 infection generates tissue-localized immunological memory in humans. *Sci Immunol*. 2021;6(65):eabl9105.
63. Turner JS, et al. Human germinal centres engage memory and naive B cells after influenza vaccination. *Nature*. 2020;586(7827):127-32.
64. Turner JS, et al. SARS-CoV-2 mRNA vaccines induce persistent human germinal centre responses. *Nature*. 2021;596(7870):109-13.
65. McInnes L, et al. *arXiv e-prints*. 2018:arXiv:1802.03426.
66. Griffiths JA, et al. Detection and removal of barcode swapping in single-cell RNA-seq data. *Nature communications*. 2018;9(1):2667.
67. McCarthy DJ, et al. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*. 2017;33(8):1179-86.
68. Gupta NT, et al. Change-O: a toolkit for analyzing large-scale B cell immunoglobulin repertoire sequencing data. *Bioinformatics*. 2015;31(20):3356-8.
69. Gadala-Maria D, et al. Automated analysis of high-throughput B-cell sequencing data reveals a high frequency of novel immunoglobulin V gene segment alleles. *Proc Natl Acad Sci U S A*. 2015;112(8):E862-70.
70. Gupta NT, et al. Hierarchical Clustering Can Identify B Cell Clones with High Confidence in Ig Repertoire Sequencing Data. *Journal of immunology (Baltimore, Md : 1950)*. 2017;198(6):2489-99.
71. Bindea G, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091-3.
72. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 2003;13(11):2498-504.

FIGURES

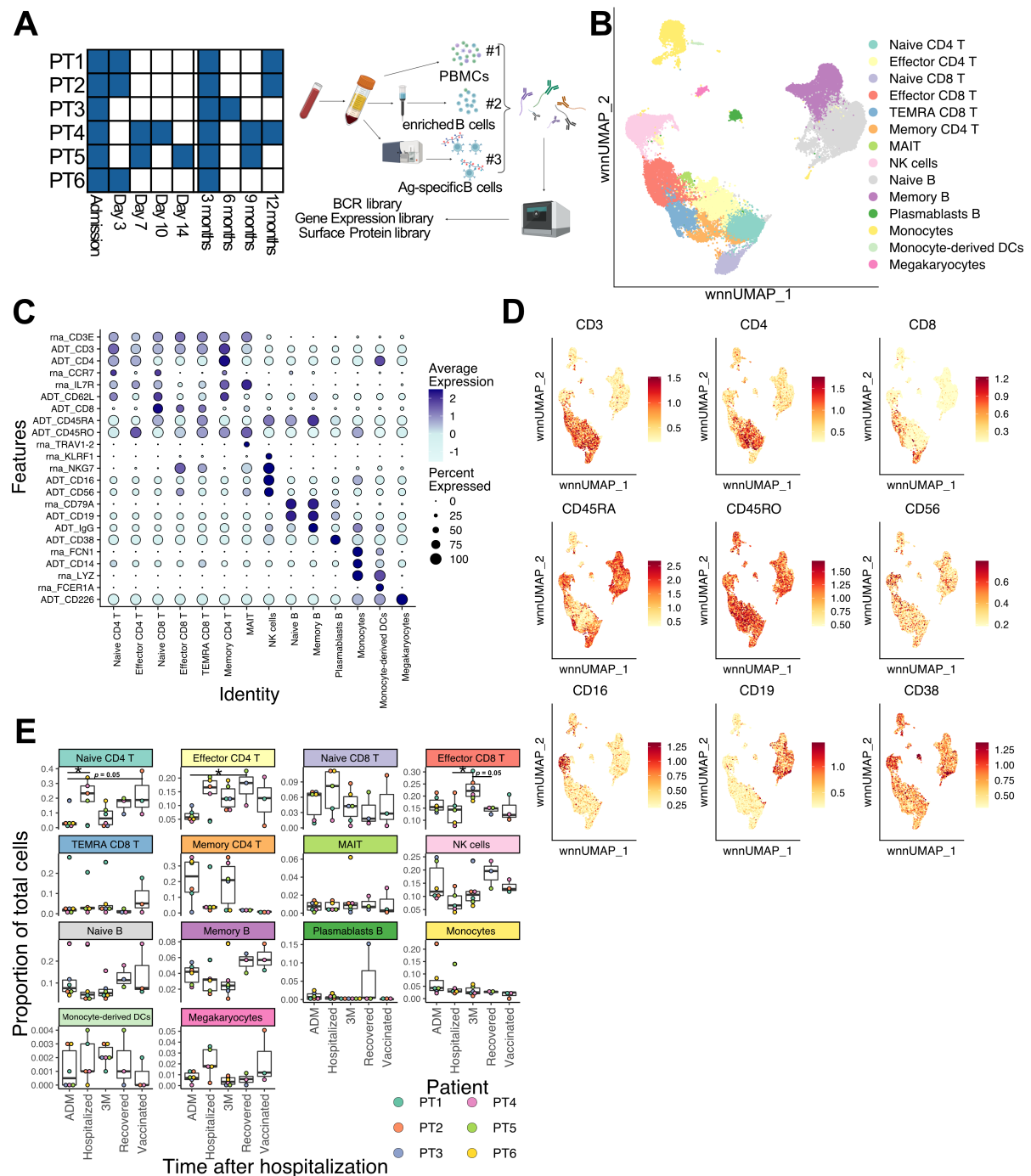


Figure 1. Longitudinal characterization of immune responses in COVID-19 patients. A) Schematic representation of sampling times, indicated in blue, and of the experimental procedure. B) UMAP plot, based on both RNA and surface protein expression, including all 25 samples analyzed. Each dot indicates an individual cell. C) Mean expression of selected marker genes or proteins. Color intensity denotes average expression, whereas dot size is the percentage of cells expressing the gene. rna before the gene name indicates gene expression while ADT indicates surface protein. D) UMAP plot showing average expression of selected proteins. Each dot is a cell and color intensity represents expression. E) Frequency for each of the identified clusters, indicated for each time of sampling and patient. Multiple comparisons were performed using one-way ANOVA with Tukey's multiple comparison test. * = $p < 0.05$

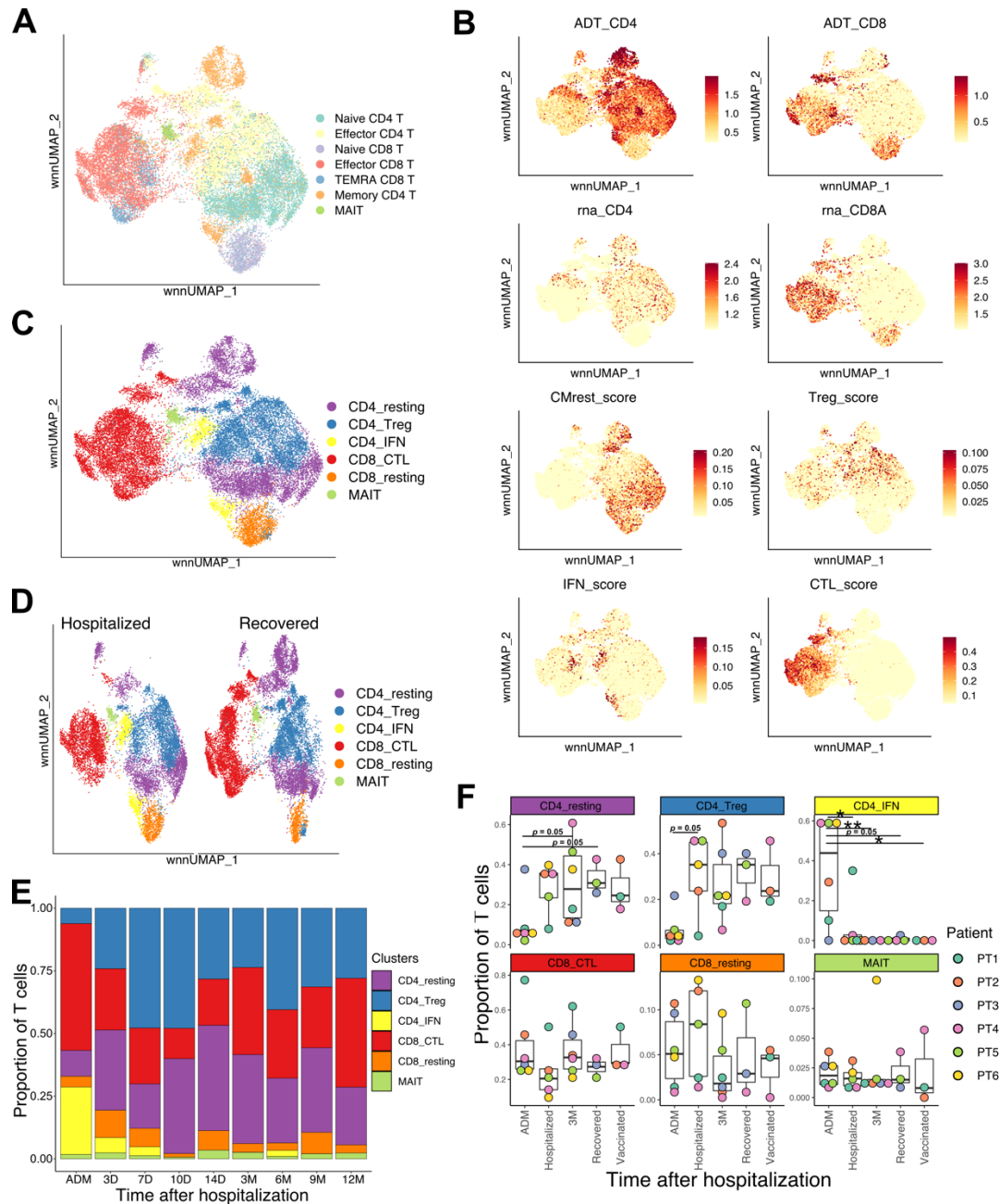


Figure 2. Longitudinal characterization of T cells in COVID-19 patients. A) UMAP plot, based on both RNA and surface protein expression, of T cells. Cluster names are based on Figure 1. Each dot indicates an individual cell. B) UMAP plot showing average expression of selected genes (rna), proteins (ADT) or combined gene signatures, according to (30). “ADT” indicates surface protein expression while “rna” shows transcript expression. “CMrest” score is genes associated with resting central memory; “Treg” score is genes associated with T regulatory cells ; “IFN” score is genes associated with IFN response while “CTL” score are genes associated with cytotoxic T cell responses. Each dot is a cell and color intensity represent expression. C) UMAP plot of T cells as in A but grouping is based on gene signature expression as in B. D) UMAP plot as in C but split based on hospitalization status. E) Quantification of the proportion of cells for each T cell cluster at each sampling time. All patients were included. F) Frequency for each of the identified clusters, indicated for each patient and time of sampling. Multiple comparisons were performed using one-way ANOVA with Tukey’s multiple comparison test. * = $p < 0.05$; ** = $p < 0.01$

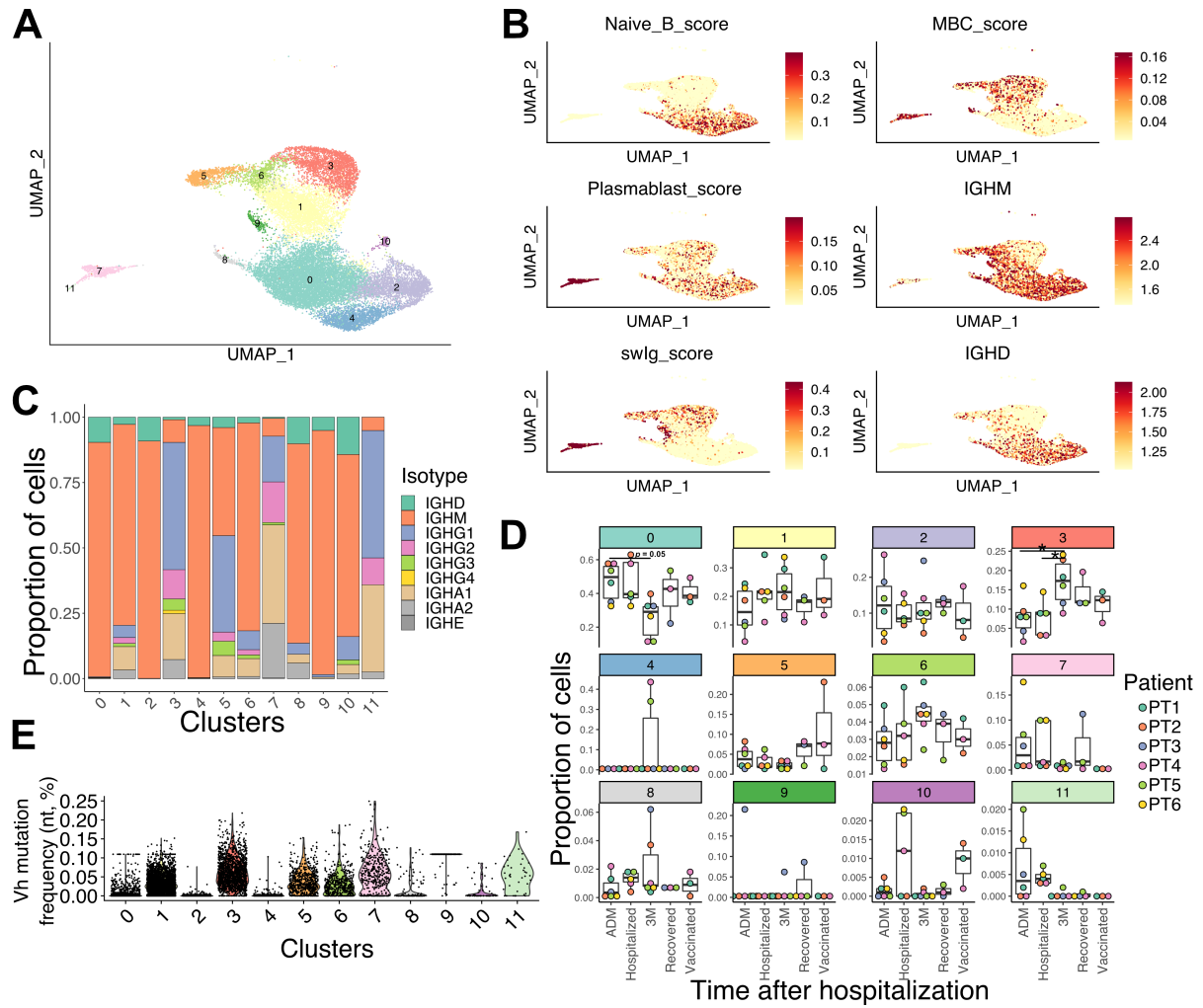


Figure 3. Longitudinal characterization of B cells in COVID-19 patients. A) UMAP plot of B cell populations. Each dot indicates an individual cell. B) UMAP plot showing average expression of selected genes or combined gene signatures, according to (9). Each dot is a cell and color intensity represent expression. C) Quantification of the proportion of cells for each B cell cluster divided by isotype. All patients were included. D) Frequency for each of the identified clusters, indicated for each time of sampling and patient. Only cells belonging to the PBMC and enriched B cell pools were considered for this analysis. Multiple comparisons were performed using one-way ANOVA with Tukey's multiple comparison test. * = $p < 0.05$. E) Graph showing Vh gene mutation frequency per UMAP clusters as in A. Each dot represents an individual cell. Multiple comparisons were performed using one-way ANOVA with Tukey's multiple comparison test. All pairwise comparisons had $p < 0.0001$, except cluster 11 vs clusters 1, 3, 5, 6 and 7, cluster 10 vs clusters 0 and 8, cluster 1 vs cluster 6, cluster 2 vs cluster 4 and cluster 3 vs cluster 7 which were all non significant.

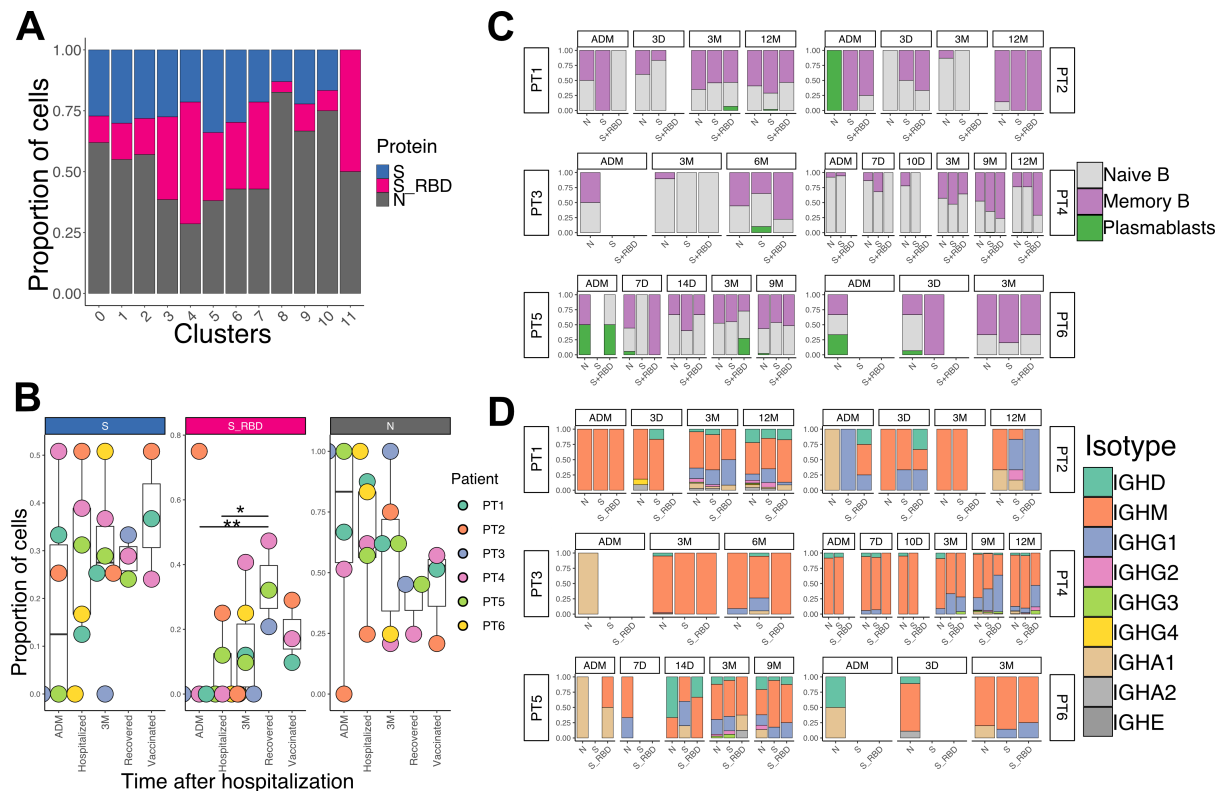
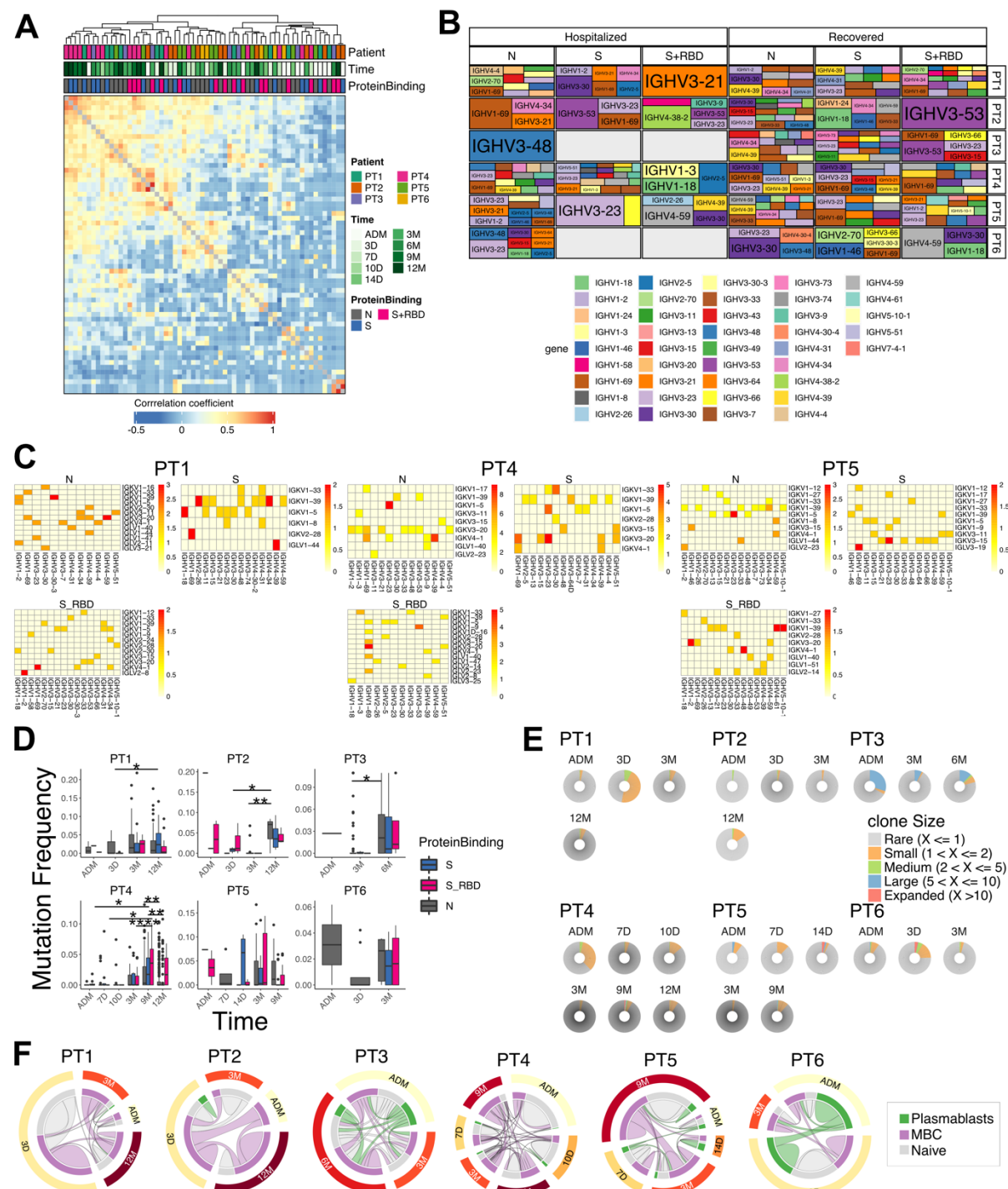


Figure 4. Longitudinal characteristics of antigen specific B cells. A) Quantification of the proportion of protein-binding cells for each B cell cluster. N is Nucleocapsid, S_RBD is receptor binding domain and Spike binders while S is Spike non-RBD binders. B) Frequency for each protein specificity, indicated for each time of sampling and patient. Multiple comparisons were performed using one-way ANOVA with Tukey's multiple comparison test. * = $p < 0.05$; ** = $p < 0.01$ C) Quantification of the proportion of B cell subtypes for each protein binding specificity, indicated for each patient and sampling time. D) Quantification of the proportion of B cell isotype for each protein binding specificity, indicated for each patient and sampling time.

Figure



5. Antigen-binding B cells are selected and persist from infection up to one year. A)

Hierarchical clustering of Pearson's correlation of the V gene repertoire. Each tile represents the correlation of the V gene repertoire. Color intensity indicates correlation strength. B) Tile plot indicating the ten most frequent V genes used per patient, protein and hospitalization status. Size of the tile is proportional to the repertoire space occupied. C) Heatmaps showing the frequency of each patient's heavy- and light-chain gene pairings for B cells binding the indicated antigens. D) Graph showing Vh gene mutation frequency indicated for each patient, time of sampling and antigen binding. Data are presented as median and interquartile range. Multiple comparisons were performed using two-way

ANOVA with Bonferroni correction for multiple comparisons. ** = $p < 0.05$, *** = $p < 0.01$, **** = $p < 0.0001$ E) Pie chart showing B cell clonal expansion indicated for each patient and time point. B cells were binned into rare clones (1 member), small (2 members), medium (between 3 and 5 members), large (between 6 and 10 members) and expanded (over 11 members). F) Circos plot showing clonal relationship within each patient at different sampling times, for clonal families with at least 2 members. Connecting lines indicate shared clones and the size of the circle and connector is proportional to the repertoire space occupied. Outer circle indicates sample time while inner circle and connectors are colored depending on cell type.

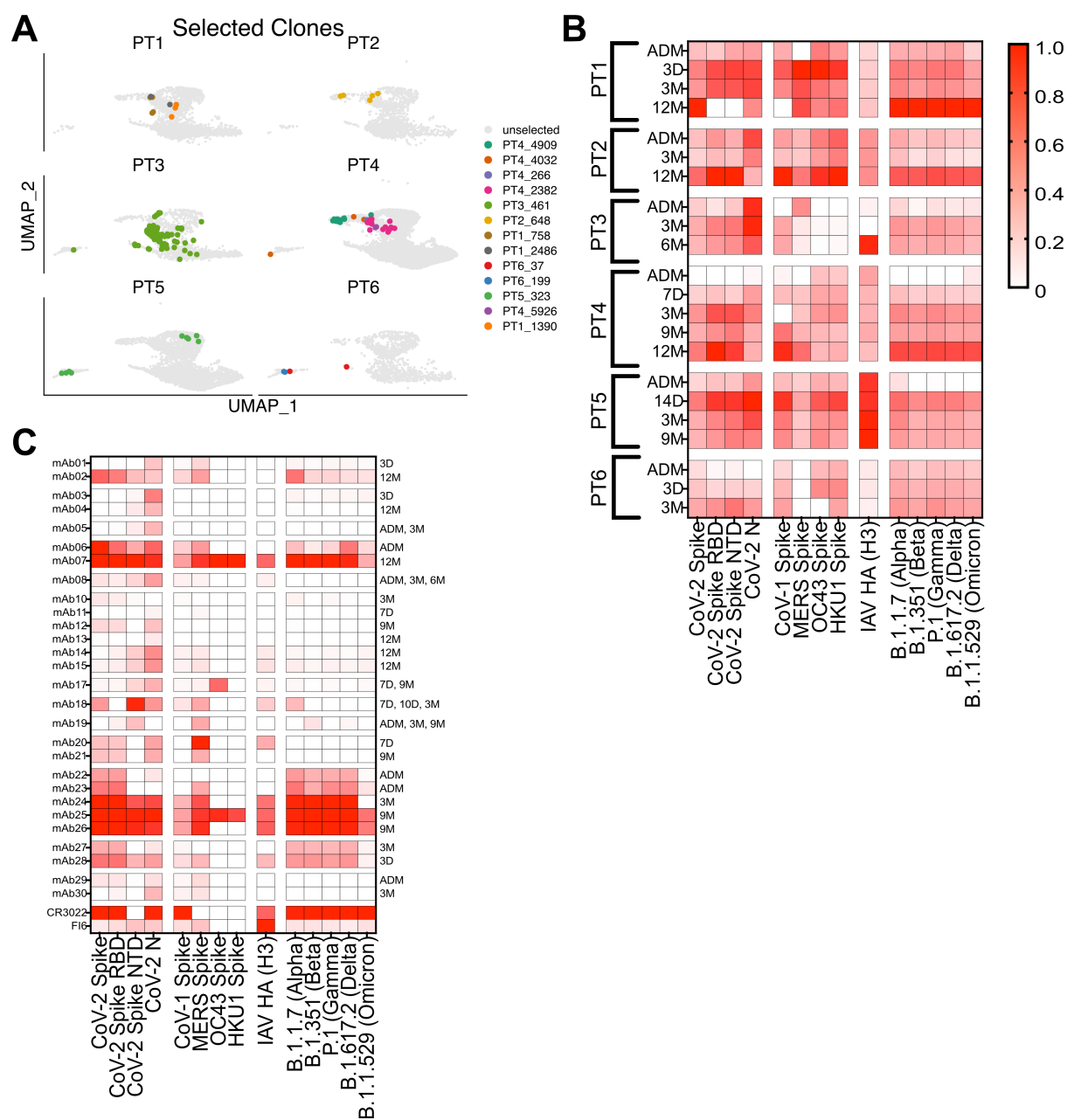


Figure 6. Continued evolution of serum antibodies and B cell responses within patients. A) UMAP plot of B cell clusters highlighting selected persistent clones per patient. Serum Ab (B) and mAb (C) binding to a panel of viral antigens as detected by an electrochemiluminescent binding assay. Color intensity indicates the normalized log binding intensity, where the top binder is set to 1.

Table 1. Clinical status of patients included in the study.

Patient	Age (at hospitalization)	Sex	Comorbidities	Admitted (days PSO ¹)	Disease severity ²	Max O ₂ requirement	Max CRP ³	Duration of hospital stay (days)	Vaccination status (at last sampling)
PT1	50	M	None	11	Severe	4L	250	5	1 dose (Moderna)
PT2	50	M	Bronchiectasis, decreased IgG2 levels, partial IgA-deficiency	7	Severe	1.5L	5	8	1 dose (Pfizer)
PT3	63	F	Asthma, hypertension	10	Severe	HFNO ⁵	220	6	unvaccinated
PT4	67	F	COPD ⁶ , psoriasis	4	Severe	6L	150	13	2 doses (AstraZeneca)
PT5	36	M	None	6	Critical	Intubated	260	20	unvaccinated
PT6	45	M	None	8	Severe	HFNO	170	5	unvaccinated

¹ post-symptom onset; ² COVID-19 Treatment Guidelines Panel of the National Institutes of Health (mild, moderate, severe, critical); ³ CRP (C-reactive protein mg/L, ref <5); ⁴ x10⁹/L (ref 0.8-4.5); ⁵HFNO (High Flow Nasal Oxygen); ⁶COPD (chronic obstructive pulmonary disease).