

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

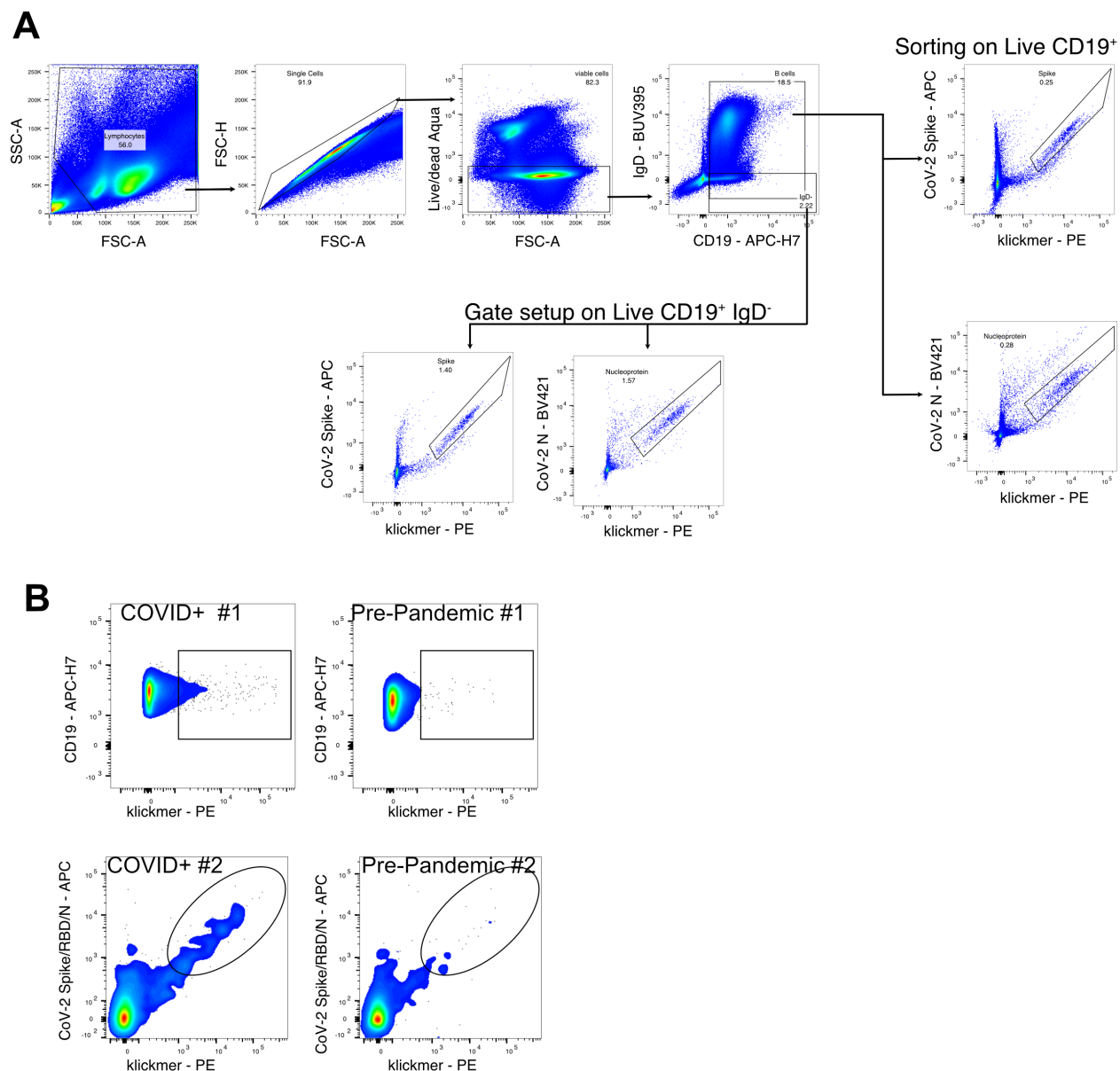
***Correspondence:**

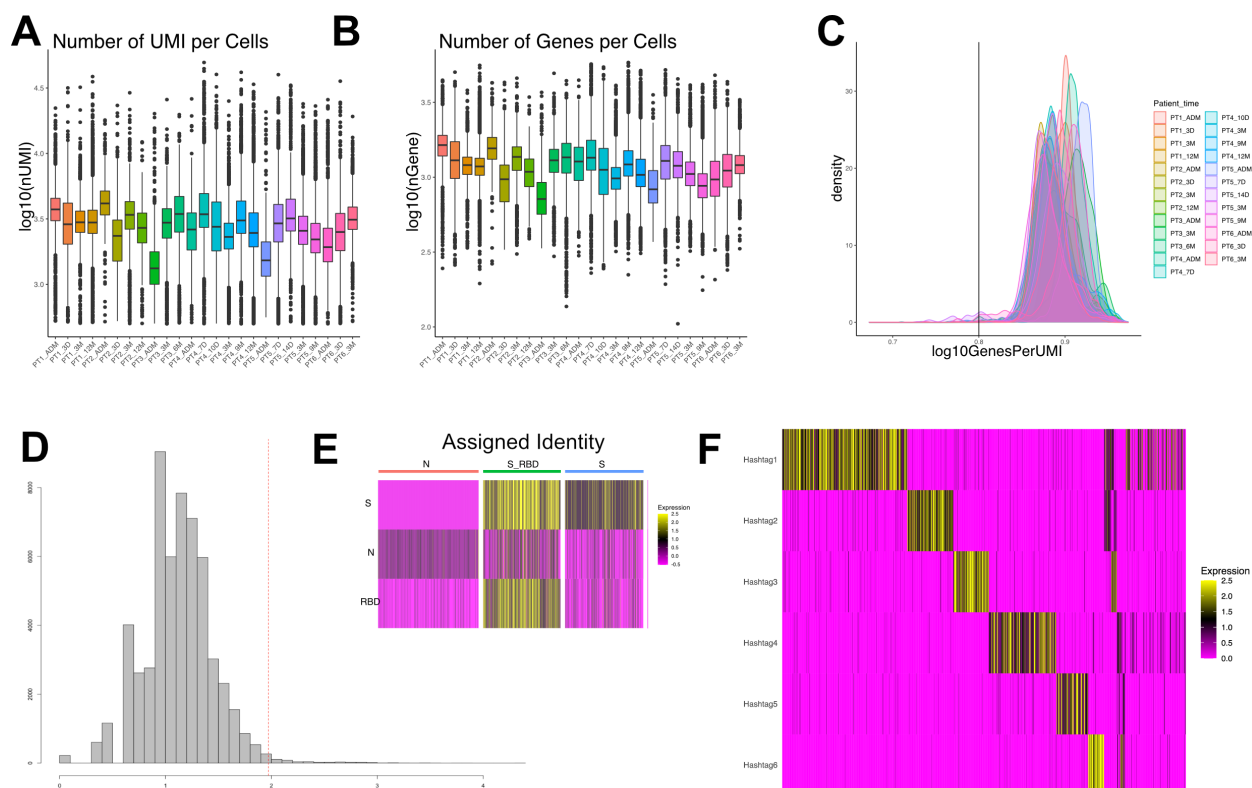
Davide Angeletti, Department of Microbiology and Immunology, Institute of Biomedicine, Medicinaregatan 7A, 413 90 Gothenburg, Sweden. Phone: +46-(0)31-786 6324. Email: davide.angeletti@gu.se

Or

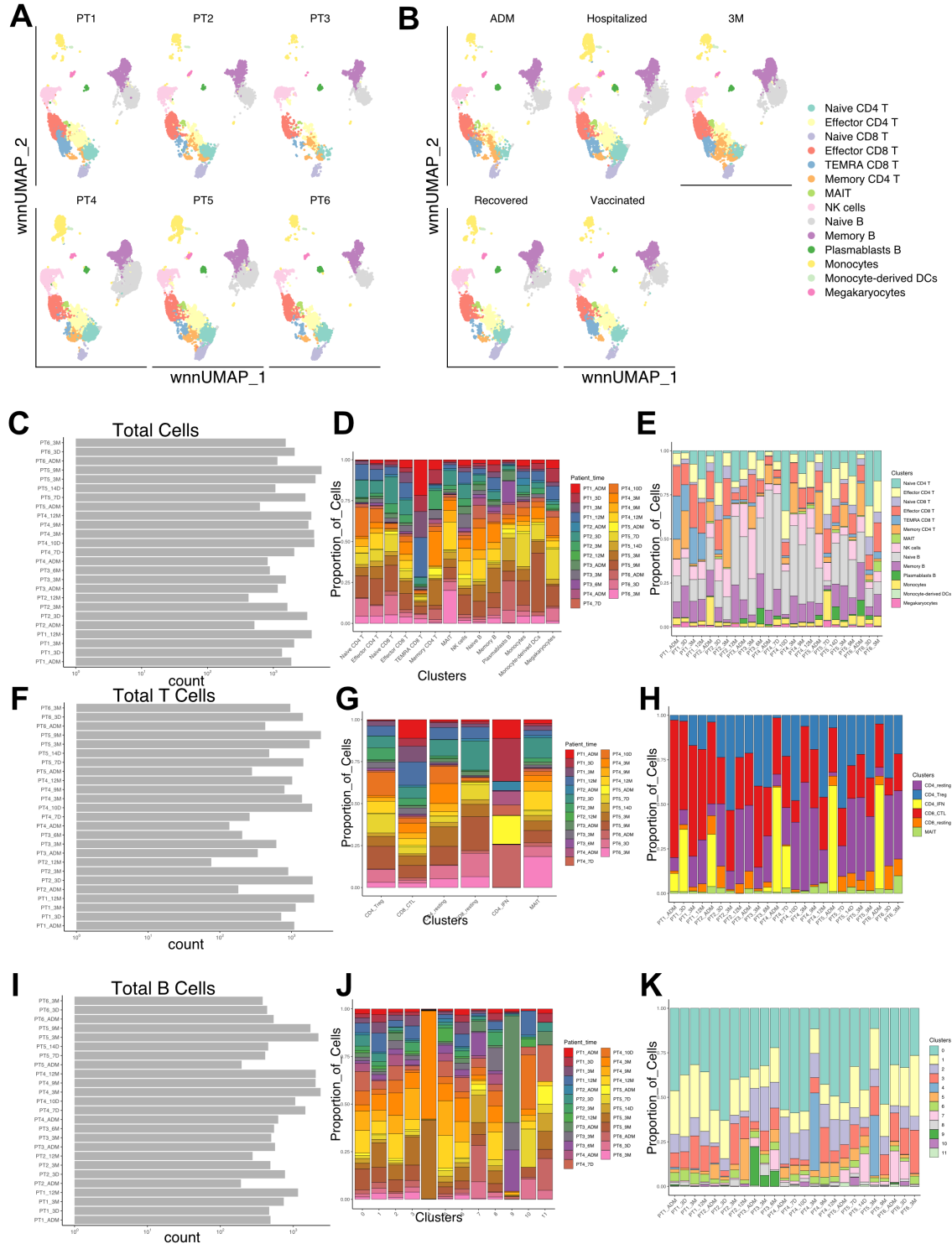
Mats Bemark, Department of Microbiology and Immunology, Institute of Biomedicine, Medicinaregatan 7A, 413 90 Gothenburg, Sweden. Phone: +46-(0)70-323 7043. Email: mats.bemark@immuno.gu.se

The authors have declared that no conflict of interest exists.

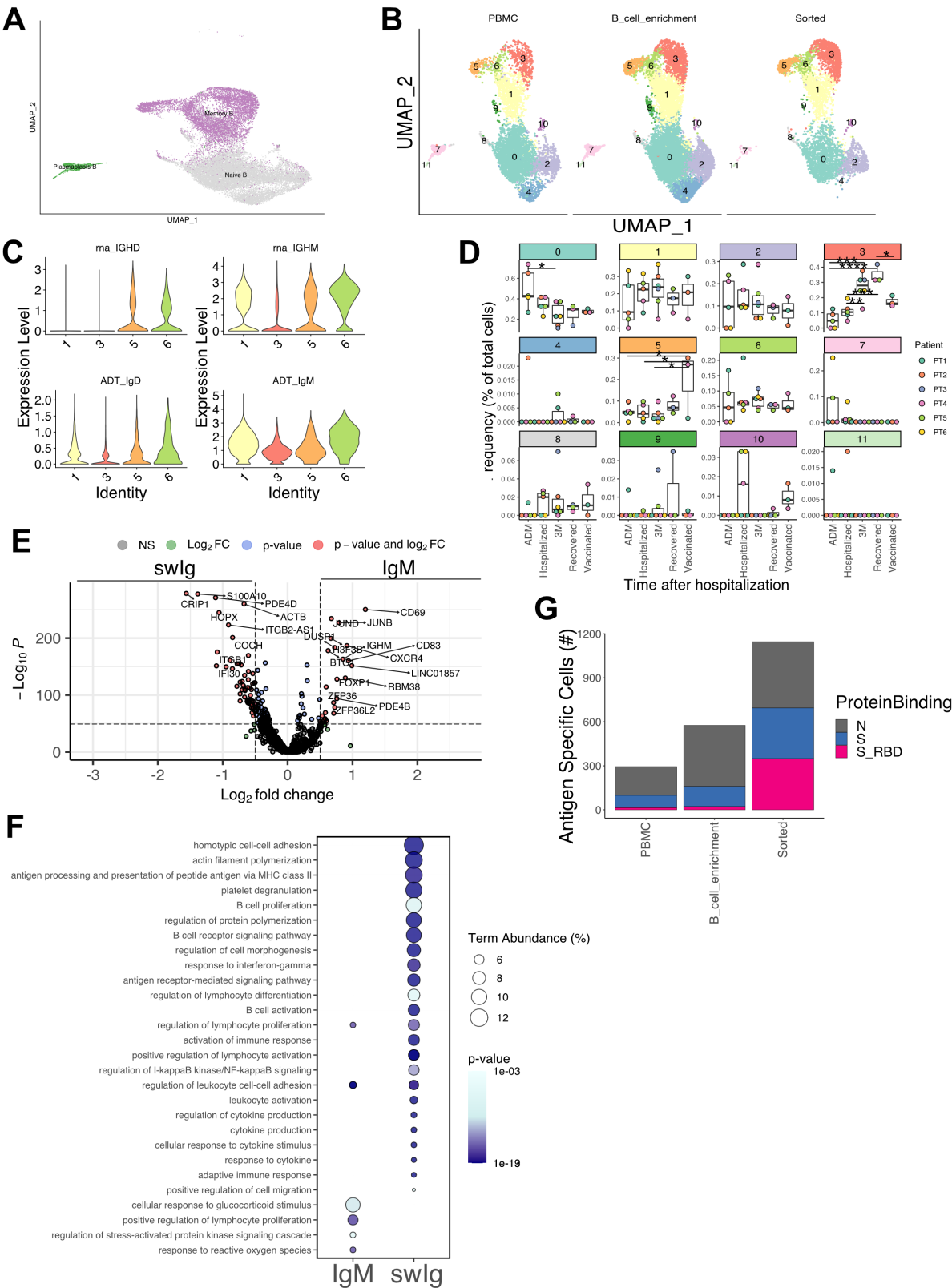




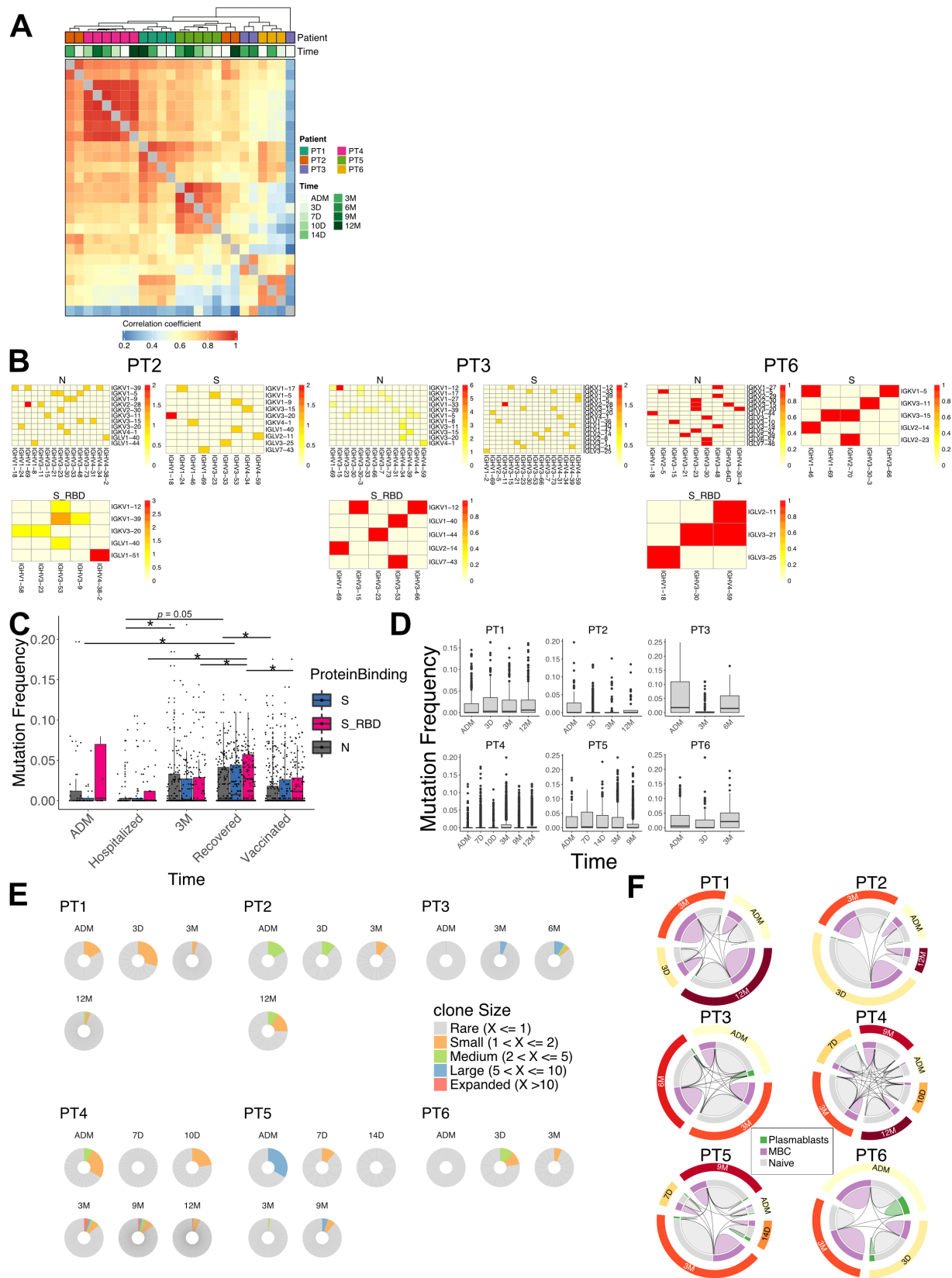
Supplementary Figure 2. Quality measures of single cell data. A) Number of UMI per cell, divided by sample. B) Number of genes per cell, divided by sample. C) Histogram showing cell complexity per sample, with the red vertical bar indicating the threshold of 0.8. D) Heatmap of CLR of Hashtags. D) Seven isotype controls were included in the CITE-Seq panel (MouseIgG1, MouseIgG2a, MouseIgG2b, RatIgG2b, RatIgG1, RatIgG2a, HamsterIgG). The coverage of these control Abs serves as a measure of non-specific binding in each cell, most notably protein aggregates. Red vertical line is the threshold for aggregates, showing very few cells (0.5% of cells). E) Heatmap of centered log ratio transformed counts (CLR) of barcoded PE-Dextramers showing specificity of the assign protein Identity. The graph considers only cells with assigned Protein Binding. F) Heatmap showing CLR of Hashtags.



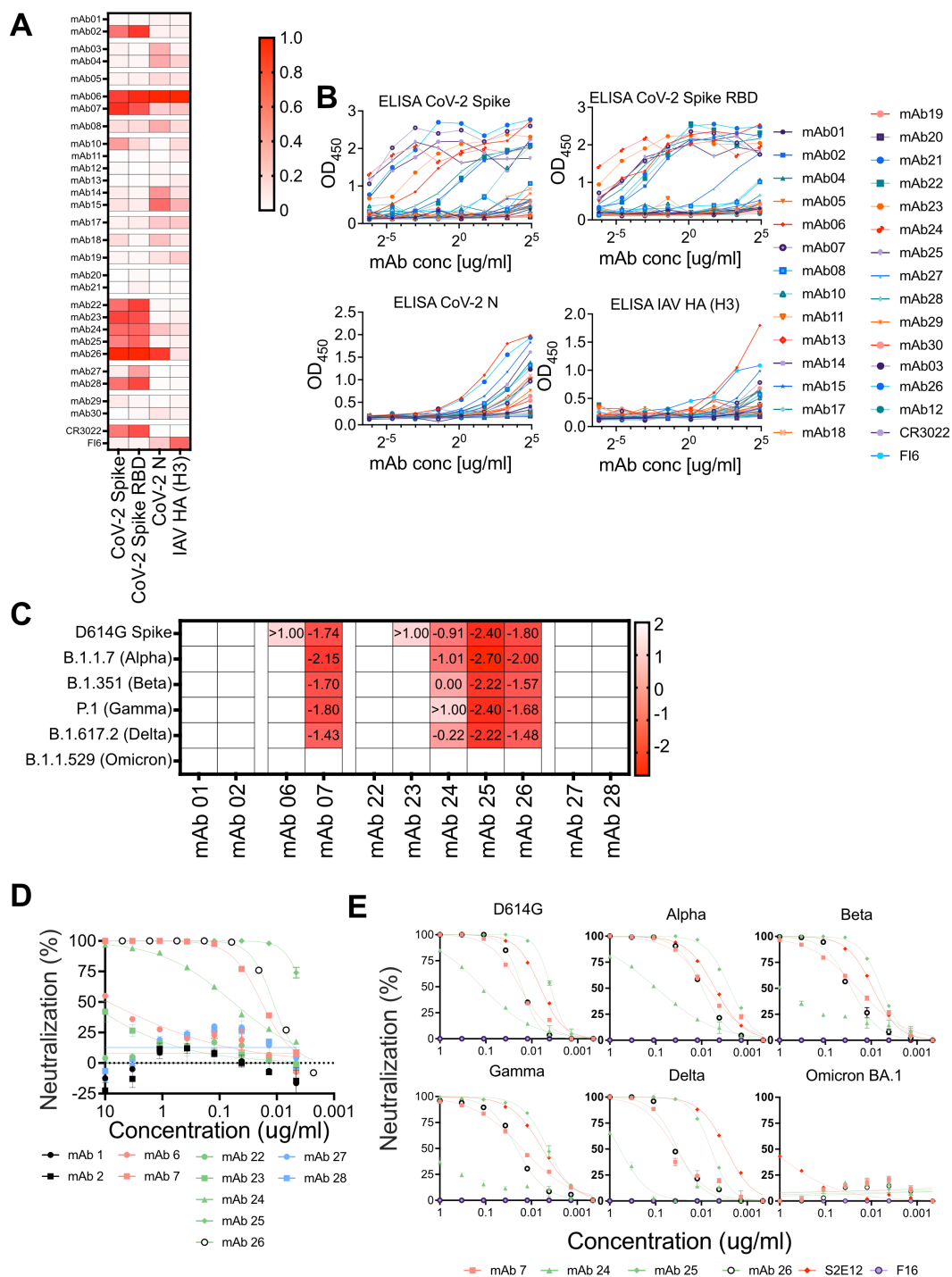
Supplementary Figure 3. A) UMAP plot as in Fig 1B but split according to patient. B) UMAP plot as in Fig 1B but split according to sampling time. Total number of cells (A), T cells (F) and B cells (I) obtained from each patient and time point. The graph reveals substantial homogeneity among total and B cells. Contribution from each sample to the total number of cells of each cluster for all cells (D), T cells (G) and B cells (J). Proportion of cells from each cluster, divided by patient and time point for all cells (E), T cells (H) and B cells (K).



Supplementary Figure 4. A) UMAP plot as in Fig 3A but with grouping defined as in Fig 1. B) UMAP plot as in Fig 3A but split according to origin of sample. C) Violin plot showing expression of IgM and IgD genes (rna) or proteins (ADT) within the MBC clusters. “ADT” indicates surface protein expression while “rna” shows transcript expression. D) Frequency for each of the identified B cell clusters, indicated for each time of sampling and patient. Only cells belonging to the antigen-sorted pool were considered for this analysis. Multiple comparisons were performed using one-way ANOVA with Tukey’s multiple comparison test. * = $p < 0.05$ E) Volcano Plot showing differential expression of genes between switched memory cells (clusters 3 and 5) and IgM-MBC (cluster 1). F) Differentially expressed genes between IgM and swIg were analyzed using ClueGo and significant pathways plotted. Size of the dot indicates the % of term in the GO category, while color intensity represents the p value. G) Number of protein binding cells for each of the populations purified. Clear enrichment is present in the sorted cell population



Supplementary Figure 5. 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) Heatmaps showing the frequency of each patient's heavy- and light-chain gene pairings for B cells binding the indicated antigens. C) Graph showing Vh gene mutation frequency for all patients combined, but divided by time of sampling and antigen binding. Data are presented as median and interquartile range. D) Graph showing Vh gene mutation frequency divided by patient and time of sampling for non-binding cells. Data are presented as median and interquartile range. Multiple comparisons were performed using two-way ANOVA with Bonferroni correction for multiple comparisons. * = $p < 0.05$ E) Pie chart showing B cell clonal expansion divided by patient and time point, for clonal families with at least 2 members. 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. 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.



Supplementary Figure 6. A) Heatmap generated by ELISA binding to CoV-2 or influenza antigens. Color intensity indicates normalized Area Under the curve (AUC). B) ELISA curves used to generate the data in A. C) mAbs neutralization of pseudotyped lentivirus expressing S proteins from SARS-CoV2 viruses with Table showing IC₅₀ of selected mAbs to viruses, as indicated on the left. Color intensity indicates the log IC₅₀ value. Value >1 indicates detectable but low neutralization while white boxes denote no neutralization. D-E) Pseudovirus neutralization curves used to generate the data in A. Error bar depicts SD of at least 3 technical replicates