



Supplementary Figure 4

Supplementary Figure 4. Major and minor SARS-CoV-2 genome variants in longitudinal samples from patients with persistent COVID-19 infection. a) Major variants in SARS-CoV-2 sequences from 17 patients with longitudinal data. Samples from the same patient are grouped together and highlighted with the same background color. White gaps indicate regions with low sequencing coverage. Major variants are computed w.r.t. the reference sequence. Vertical lines indicate the location of the variants in the SARS-CoV-2 genome. The height of the line indicates the frequency of the variant in the viral population. b) dN/dS ratio across all SARS-CoV-2 protein products (ORF1ab, ORF1a, S, ORF3a, E, M, ORF6, ORF7a, ORF7b, ORF8, N, ORF10) for all samples with available sequencing), organized by immunological clusters described in Figure 2b. P values as indicated c) Correlation of viral sequencing depth with viral entropy (Pearson $r = -0.57$, p-value $\ll 1e-4$; Spearman $r = -0.55$, p-value $\ll 1e-4$).