Component	Reference range	Day 49	Day 89
Lymphocyte count	1.00 – 2.80 X 10^9/l	0.16	1.97
CD3 %	%	49	22
CD3 total	0.70 – 2.10 X 10^9/l	0.08	0.43
CD4 %	%	29	11
CD4 total	0.30 – 1.40 X 10^9/l	0.05	0.22
CD8 %	%	21	10
CD8 total	0.20 – 0.90 X 10^9/l	0.03	0.2
CD19 %	%	21	58
CD19 total	0.10 – 0.50 X 10^9/l	0.03	1.14
CD56 %	%	23	19
CD56 total	0.09 – 0.60 X 10^9/l	0.04	0.37
IgG	6.00 – 16.00 g/l	<1.40	1.9
IgA	0.80 – 4.00 g/l	<0.30	<0.30
IgM	0.50 – 2.00 g/l	1.27	1.92
Sars-CoV-2 Total antibody	(Siemens)	Negative	Positive

Supplemental Table 1: Immunological parameters of patient pre- and post-receipt of convalescent plasma (CP) demonstrating profound lymphopenia with reduction in total serum IgG and IgA. The patient continued to have normal serum IgM which either reflects his underlying lymphoproliferative state or ongoing infection. Following receipt of CP the presence of SARS-CoV-2 antibodies is evident.

Supplementary Table 2. Characteristics of genomes used in the phylogenetic analysis in Figure 2B. Both NextStrain and pangolin lineages were identified using the NextClade server v0.8 and Pangolin COVID-19 Lineage Assigner v2.0.8 (https://pangolin.cog-uk.io/). Mean mutation rate was calculated using BEAST v2.6.3 using a strict clock, HKY and coalescent exponential population. From this the number of mutations per year was calculated by multiplying the mutation rate per site by the SARS-CoV-2 genomes length (29903).

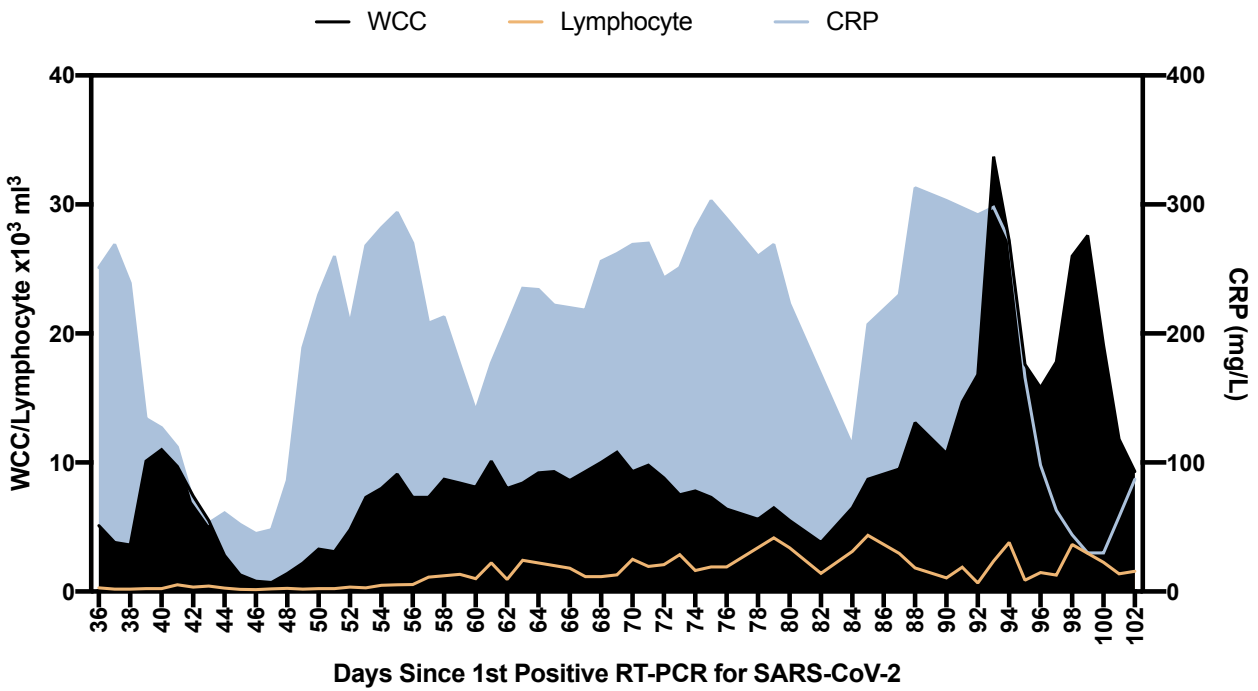
Patient	Clinical	NextStrain Clade	Pangolin Lineage	Mean mutation rate per site	Error of Mean	Mutations per year
Global GSAID Average	-	-	-	7.30E-04	-	21.8
Choi et al	Antiphospholipid syndrome	20C	B.1	1.41E-03	4.90E-06	42.3
Avanzato et al	Chronic lymphocytic leukemia and HLA	19B	A.1	1.33E-04	3.24E-05	4.0
Control patient 1 (purple)	X-linked Agammaglobulinemia	19A	B.2.6	1.93E-04	3.87E-05	5.8
Control Patient 2 (blue)	Chronic renal impairment	20B	B.1.1	2.47E-04	8.43E-06	7.4
Control Patient 3 (red)	Heart disease	20B	B.1.1.35	5.84E-04	2.76E-04	17.4
Study patient	Marginal B Cell Lymphoma	20B	B.1.1.35 & B.1.1.1	1.00E-03	1.06E-05	29.9

Supplementary Table 3. Prevalence of selected Spike glycoprotein mutations at sequential time points and sequencing depth (number of reads covering the amino acid position) measured by both short-read (Illumina) and long-read (Oxford Nanopore) methods. There was low coverage of S:P330S and S:D796H in the final three timepoints, measured by both short- and long-read methods. *Based on the binomial probability distribution and using the Illumina sequencing error value of 0.1%, a minimum depth of 1,000x and a minimum number of 20 reads is sufficient to call minority variants based on a binomial probability distribution. **Low Read depth nanopore sequences were re-sequenced using the Illumina platform to confirm the presence of variants of interest.

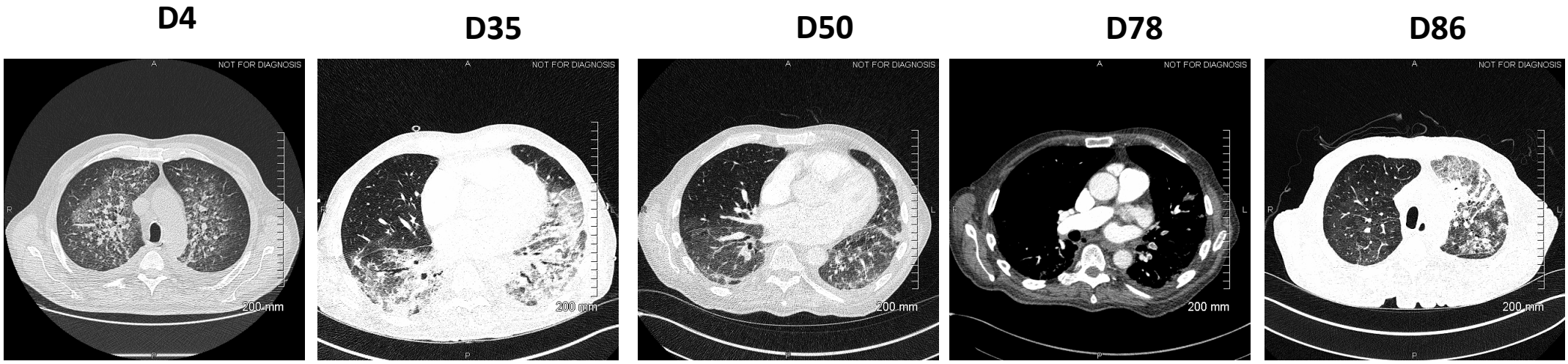
	S:W64G Nanopore		S:W64G Illumina		S:**Δ**H69/V70 Nanopore		S:**Δ**H69/V70 Illumina		S:P330S Nanopore		S:P330S Illumina		S:D796H Nanopore		S:D796H Illumina	
Days since 1st positive RT-PCR	%	Read depth	%	Read depth	%	Read depth	%	Read depth	%	Read depth	%	Read depth	%	Read depth	%	Read depth
1	0.4	1200	0.0	1079	8.4	1263	0.0	1092	1.1	888	0.0	1274	0.1	1360	0	2301
37	0.8	3475	0.0	1424	0.1	3475	0.6	1433	0.8	3319	0.0	857	0.1	4131	0.1	1342
45	0.7	300	0.0	994	0.0	300	0.0	997	0.7	164**	0.0	1327	0.0	360	0.0	2204
54	0.4	5155	0.0	2085	0.1	5157	0.0	2105	1.6	4591	0.1	1318	2.9	7897	0.2	1704
66	0.0	6105	0.0	2019	4.4	6099	0.0	1688	9.0	3395	0.0	2042	0.1	7522	0.0	3050
82	0.7	9920	0.0	2182	93.0	5430	99.0	1632	0.9	5389	0.0	1656	95.0	12825	96.7	3050
86	0.8	7393	0.0	1840	6.9	7652	0.0	1734	0.9	4855	0.0	1748	2.8	9289	6.0	1737
89	0.5	4028	0.0	1867	1.0	3870	6.0	1652	0.9	1283	0.0	2043	8.3	4212	8.3	1394
93	80.2	6105	96.5	1933	9.2	6537	0.0	1627	92.8	3509	96.9	1768	0.2	3799	0.0	1552
95	60.3	8466	62.7	1958	10.4	8795	12.0	1747	73.3	6103	78.3	1687	6.3	14702	5.9	1374
98	0.3	4997	0.0	1805	93.5	2751	82.6	1539	1.3	2597	0.0	1914	78.0	4212	78.9	1496
99	3.6	4772	0.0	1960	90.0	2117	67.7	5111	0.0	33**	20.0	50*	55.7	7456	57.9	56*
100	3.1	4767	0.0	2114	96.5	5036	70.3	1258	4.8	21**	0.0	30*	68.4	613	70.1	30*
101	0.7	696	0.0	1896	91.0	384	86.3	2054	0.0	75**	0.0	520	78.4	37**	96.0	175

Supplementary Table 4. Global prevalence of selected spike mutations detailed in this paper. All high coverage sequences were downloaded from the GISAID database on 11th November and aligned using MAFFT. The global prevalence of each of the six spike mutations W64G, ΔH69/V70, Y200H, T240I, P330S and D796H were assessed by viewing the multiple sequence alignment in AliView, sorting by the column of interest, and counting the number of mutations.

Mutation	Number of Sequences	Global Prevalence
W64G	0	0.00%
ΔH69/V70	2920	1.38%
Y200H	7	0.00%
T240I	34	0.02%
P330S	130	0.06%
D796H	25	0.01%

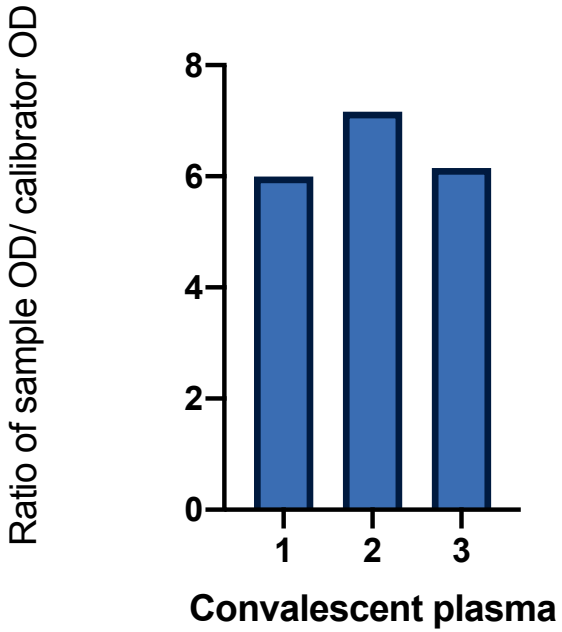


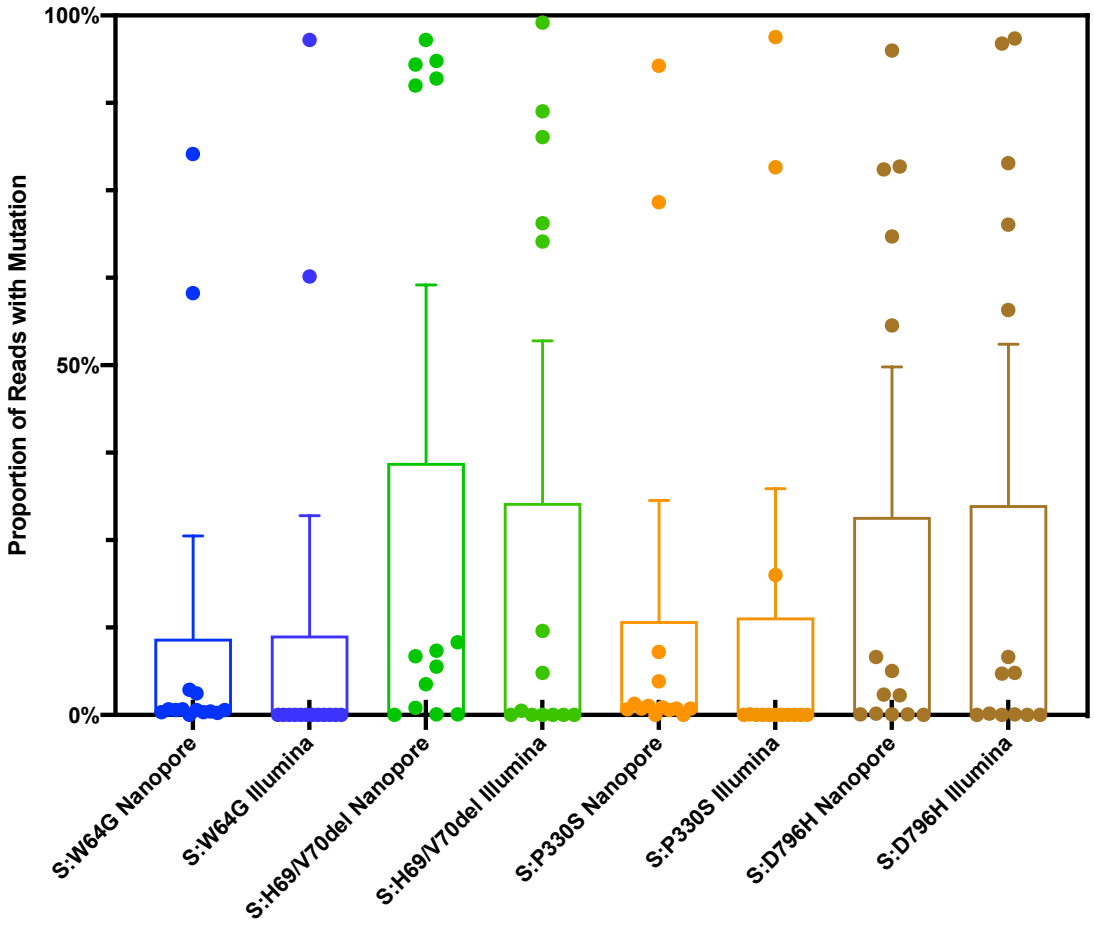
Supplementary Figure 1: Blood parameters over time in patient case: White cell count (WCC) and lymphocyte counts are expressed as $x10^3$ Cells/mm^3. CRP: C reactive protein



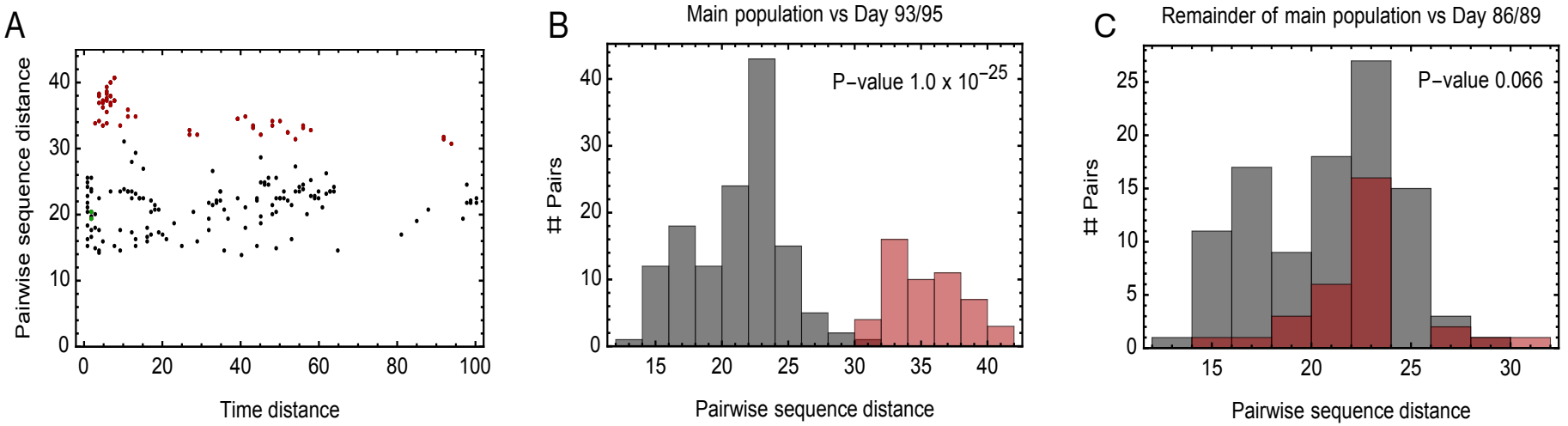
Supplementary Figure 2: Serial CT images following detection of SARS-CoV-2.
The patient initially presented with ground glass and peribroncho-vascular consolidation with associated intralobular septal thickening/reticulation and architectural distortion and interlobular septal thickening. By day 50 there is some improvement with evidence of resolving pnuemonitis, however, his condition deteriorated following the detection of bilateral pulmonary emboli, a well recognized complication of SARS-CoV-2. Despite multiple therapeutic interventions, the patient's condition deteriorated with worsening of inflammatory changes and chronic organizing pneumonia (COP), particularly on the left, and ongoing changes compatible with persistent SARS-CoV2 infection.

Supplementary Figure 3: SARS-CoV-2 antibody titres in convalescent plasma. Measurement of SARS-CoV-2 specific IgG antibody titres in three units of convalescent plasma (CP) by Euroimmun assay.



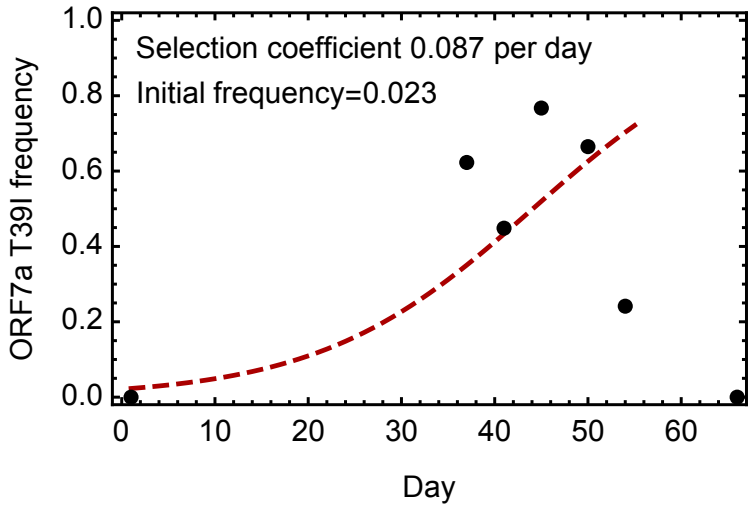


Supplementary Figure 4. Concordance between short-read (Illumina) and long-read (Oxford Nanopore) sequencing methods for ten samples. Points are the frequency of the variant at each timepoint. Error bars are 95% CI. There was good concordance for the four measured spike mutations between the two methods, and no significant difference between the proportion of reads measured by both methods.



Supplementary Figure 5: Additional evidence for within-host cladal structure. A. Pairwise distances between samples measured using the all-locus distance metric plotted against pairwise distances in time (measured in days) between samples being collected. Internal distances between samples in the proposed main clade are shown in black, distances between samples in the main clade and samples collected on days 93 and 95 are shown in red, and internal distances between samples collected on days 93 and 95 are shown in green. **B.** Pairwise distances between samples in the larger clade (black) and between these samples and those collected on days 93 and 95 (red). The median values of the distributions of these values are significantly different according to a Mann Whitney test. **C.** Pairwise distances between samples in the main clade, once those collected on days 86, 89, 93, 95 have been removed (black) and between these samples and those collected on days 86 and 89 (red). The median values of the distributions of these values are not significantly different at the 5% level according to a Mann Whitney test.

Supplementary Figure 6: Assessment of possible selection acting upon viral sequence variants. A deterministic model of selection acting at a single locus (red) was fitted to allele frequency data from samples collected from the patient (black circles). Plot shows the changes in frequency of the variant T39I in ORF7a during the earliest part of the infection. The loss of the variant is consistent with clonal interference with other viruses in the population.



Supplementary Figure 7. Maximum-Likelihood global phylogeny of SARS-CoV-2. To ensure that the patient case was not resultant of a superinfection event, all patient sequences were aligned with a snapshot of global SARS-CoV-2 sequences downloaded from the GISAID database between 1st April and 5th May (26472 sequences, only full sequences and excluding all low coverage sequences). Sequences were aligned using MAFFT and a maximum-likelihood tree inferred with IQTREE v2.1.3. All 23 sequenced from the patient case (red) formed a distinct clade, suggesting that all viral populations diversified within-host.

Supplementary Figure 8. Global spike-only phylogeny of SARS-CoV-2 sequences carrying the Spike ΔH69/V70 double deletion. All sequences carrying the double-deletion (red) were downloaded from the GISAID database and aligned with 8000 globally representative spike sequences (black). Sequences with ΔH69/V70 (red, n=2920) formed a separate clade, with some sequences interspersed throughout the tree. More than 80% of sequences with the double-deletion also carried the N439K mutation.