

**Title: Remdesivir induced viral RNA and subgenomic RNA suppression, and evolution of viral variants in SARS-CoV-2 infected patients**

**Summary Sentence:** Deep sequencing of longitudinal samples from SARS-CoV-2 infected paediatric patients identifies evidence of remdesivir-associated inhibition of viral replication *in vivo* and uncovers evidence of within host evolution of distinct viral genotypes.

**Authors:** Florencia A.T. Boshier<sup>\*1</sup>, Juanita Pang<sup>\*1,2</sup>, Justin Penner<sup>3</sup>, Joseph Hughes<sup>4</sup>, Matthew Parker<sup>5</sup>, James Shepherd<sup>4</sup>, Nele Alders<sup>3</sup>, Alasdair Bamford<sup>3</sup>, Louis Grandjean<sup>1</sup>, Stephanie Grunewald<sup>6</sup>, James Hatcher<sup>7</sup>, Timothy Best<sup>7</sup>, Caroline Dalton<sup>8</sup>, Patricia Dyal Bynoe<sup>1</sup>, Claire Frauenfelder<sup>9,10</sup>, Jutta Köeglmeier<sup>11</sup>, Phoebe Myerson<sup>1</sup>, Sunando Roy<sup>1</sup>, Rachel Williams<sup>1</sup>, The COVID-19 Genomics UK (COG-UK) consortium<sup>12\*\*</sup>, Emma C Thomson<sup>4</sup>, Thushan I de Silva<sup>5</sup>, Richard A. Goldstein<sup>2</sup>, Judith Breuer<sup>†1,7</sup>

\*These authors contributed equally to this work as co-first authors.

\*\*Full list of consortium names and affiliations can be found at <https://www.cogconsortium.uk/about/>

† To whom correspondence should be addressed. Email: [j.breuer@ucl.ac.uk](mailto:j.breuer@ucl.ac.uk)

**Affiliations:**

<sup>1</sup>Department of Infection, Immunity and Inflammation, UCL Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

<sup>2</sup>Division of Infection and Immunity, University College London, London, United Kingdom.

<sup>3</sup>Department of Infectious Disease, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom.

<sup>4</sup>MRC-University of Glasgow Centre for Virus Research, Glasgow, United Kingdom.

**NOTE:** This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

<sup>5</sup>Department of Infection, Immunity and Cardiovascular Diseases, The Florey Institute, University of Sheffield, Sheffield, United Kingdom.

<sup>6</sup>Department of Metabolic Medicine, UCL Great Ormond Street Institute of Child Health, London, London United Kingdom.

<sup>7</sup>Department of Microbiology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom.

<sup>8</sup>Department of Pharmacy, Great Ormond Street Hospital for Children NHS Trust, London, UK

<sup>9</sup>Department of Ears, Nose and Throat, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom.

<sup>10</sup>Division of Surgery, University of Adelaide, South Australia

<sup>11</sup>Department of Gastroenterology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom.

<sup>12</sup> <https://www.cogconsortium.uk>

## Abstract

While changes in SARS-CoV-2 viral load over time have been documented, detailed information on the impact of remdesivir and how it might alter intra-host viral evolution is limited. Sequential viral loads and deep sequencing of SARS-CoV-2 recovered from the upper respiratory tract of hospitalised children revealed that remdesivir treatment suppressed viral RNA levels in one patient but not in a second infected with an identical strain. Evidence of drug resistance to explain this difference was not found. Reduced levels of subgenomic (sg) RNA during treatment of the second patient, suggest an additional effect of remdesivir on viral replication that is independent of viral RNA levels. Haplotype reconstruction uncovered persistent SARS-CoV-2 variant genotypes in four patients. We conclude that these are likely to have arisen from within-host evolution, and not co-transmission, although superinfection cannot be excluded in one case. Sample-to-sample heterogeneity in the abundances of variant genotypes is best explained by the presence of discrete viral populations in the lung with incomplete population sampling in diagnostic swabs. Such compartmentalisation is well described in serious lung infections caused by influenza and *Mycobacterium tuberculosis* and has been associated with poor drug penetration, suboptimal treatment and drug resistance. Our data provide evidence that remdesivir is able to suppress SARS-CoV-2 replication *in vivo* but that its efficacy may be compromised by factors reducing penetration into the lung. Based on data from influenza and *Mycobacterium tuberculosis* lung infections we conclude that early use of remdesivir combined with other agents should now be evaluated.

## Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which causes Coronavirus Disease 2019 (COVID-19), was first identified in Wuhan, China in December 2019. On 11th March 2020, the WHO declared COVID-19 a global pandemic (1). Since then, an estimated 50 million people have been infected of whom up to 2.5% have died (2). A number of studies have assessed nasal and oropharyngeal viral load data from longitudinally sampled SARS-CoV-2 infected patients. Their findings reveal wide variations in viral load at presentation (3–6). However, milder disease and clinical recovery are associated with lower and declining viral load respectively, pointing to its potential use as a biomarker for antiviral drug response. Remdesivir, an RNA-dependent-RNA polymerase (RdRp) inhibitor, has been shown in one large randomised clinical trial (RCT) to be effective against SARS-CoV-2, although another large study, showed no clinical benefit and smaller studies have shown limited or no impact on clinical recovery (3,7–11). Where clinical trial data are lacking or contradictory we have previously used deep pathogen sequencing, mutational analysis and evolutionary modelling to gain insight into the impact of repurposed drugs, including RdRp inhibitors similar to remdesivir, on serious respiratory RNA viral infections in hospitalised patients (12,13). These studies have revealed drug-related mutational signatures, evidence of viral compartmentalisation in the lung and previously unrecognised synergy between combination therapies associated with changes in viral loads and improved clinical outcomes. Here we report the application of similar methods in a personalised-medicine-approach to investigating the impact of remdesivir on SARS-CoV-2 within an individual and to evaluate potential biomarkers that can be used to monitor clinical efficacy. The data provide further insights into fundamental questions of SARS-CoV-2 evolution and co-infection.

## Results

### *Overview of Patients*

We analysed nine hospitalised SARS-CoV-2 positive paediatric cases with a mean age at time of infection of 4.7 years old (range 0-14 years old), who were repeatedly sampled during the course of their SARS-CoV-2 infection. A summary of their clinical features is shown in Table 1. Five patients had pre-existing co-morbidities associated with primary or secondary immunodeficiency (Table 1). Of the nine patients, four (A, D, G and H), of whom only H was known to be immunocompromised, were admitted to the Paediatric Intensive Care Unit (PICU) (supplementary figure 1). The remaining patients were cared for in appropriate source isolation and were not co-located during their SARS-CoV-2 infection (supplementary figure 1). Patients, A, D, and G received 8 - 10 days of remdesivir through the compassionate access programme. Patient A and G received 200 mg loading dose followed by 100 mg daily and Patient D received 5mg/kg (10 mg) loading dose and 1.25 mg/kg (2.5 mg) once daily (Table 1).

### *Viral load trajectories by cycle threshold and clinical markers of infection*

Figure 1 shows the viral polymerase chain reaction (PCR) cycle threshold (ct) values for all nine patients for forty days following their first positive nasopharyngeal aspirate (NPA) available to us. Viral RNA was measured in NPAs for all patients and/or upper airway secretions for those who were intubated (patients A, D, G, and H). Patients A also had a Bronchoalveolar lavage. In agreement with earlier studies (14), the ct values showed considerable day-to-day variation of between 0.16 and 14.4 cycle numbers (median 5.5 cycle numbers). Viral RNA continued to be detectable for 7 to over 50 days (median 16 days) following the first positive sample (supplementary figure 2). Of the three patients who received remdesivir, only patient D had total suppression of viral RNA during treatment followed by rebound of virus after treatment cessation (figure 1). The four ICU patients were clinically most

unwell, requiring assisted ventilation. All three remdesivir-treated patients showed clinical improvement after starting the drug, associated with falls in temperature (all) and inflammatory markers (A and G) (supplementary figure 3). All three were weaned from conventional ventilation before the treatment course was completed with decreases in oxygen requirements. In patient D, a significant reversal of respiratory deterioration was noted once remdesivir was started; inhaled nitric oxide was stopped within 96 hours coincident with weaning from high frequency oscillatory to conventional ventilation. Patient D, who alone required inotropic support, achieved haemodynamic stability off inotropes within 5 days of starting remdesivir.

#### *Inter and intra-host phylodynamics of SARS-CoV-2*

To investigate the possibility of remdesivir resistance in patients A and G, neither of whose viral RNA was suppressed by remdesivir treatment (figure 1), and in patient D whose virus rebounded following cessation of treatment, we deep sequenced all samples as well as those from untreated patients for comparison. The sequencing metrics for the samples are summarised in Supplementary Table 1. Relative to their first available sample, there were nine polymorphisms identified in viruses from patients A, H and I of which five were non-synonymous with four in the ORF1ab (nsp3,4,5) and one in the Spike protein, S2 subdomain (supplementary table 2). This latter mutation P812L in patient I is located in a T-cell epitope (15). None of the identified SNPs were sites identified as common homoplasies or those known to be susceptible to Illumina sequencing error, none have been associated with remdesivir resistance and none other than P812L were in known or predicted immune epitopes (16–22). Both consensus level mutations in patient A, who received remdesivir were non-synonymous. One was located in the nsp4 (responsible for formation of double membrane vesicle) and the other in nsp5 (a chymotrypsin-like protease responsible for polypeptides cleaving at 11 sites downstream to release nsp4-nsp16 and inhibition of interferon signalling)

of ORF1ab (18,23,24). Nsp5 has been shown to mediate the maturation of nsps downstream which is an essential part of the viral lifecycle (23); this has made it a desirable target for antiviral drug production (25,26). However, the nsp5 mutation P241L observed in patient A was not found to be in an active site. Virus from patient D, which rebounded following cessation of remdesivir treatment, was genetically identical to pre-treatment virus.

Patients H and I who did not receive remdesivir, also had consensus level sequence changes (supplementary table 2). Only one of four mutations for patient H was non-synonymous and this was in nsp3 (a papain-like protease) in ORF1ab. Two of three mutations in patient I were non-synonymous, one in nsp1 (responsible for cellular mRNA degradation and inhibition of interferon signalling) in ORF1ab and the other upstream of the furin cleavage site and the S2' cleavage site in the spike protein (27,28). This Spike mutation although not predicted to alter structure, does lie within a known CD4 epitope although it is not known whether it would abrogate T cell binding (supplementary figure 4) (15). Patient E was negative for SARS-CoV-2 on consecutive samples obtained on days 16 and 22. Virus detected again at days 31 and 35 was identical to all other sequences from this patient. We observed changes in the consensus sequences between different samples from patients A, H and I. Sequences A at time point 6, 7, and 8 were identical to those of samples sequenced from patient D as depicted in Figure 2 A. No evidence of laboratory contamination to explain the identity between A and D sequences was found, with samples from these patients being sequenced in separate batches on separate days.

#### *No mutagenic signature identified for remdesivir*

Ribavirin and favipiravir, which like remdesivir are nucleoside analogues, have been shown to inhibit RdRps at higher drug concentrations while at lower concentrations they appear to drive the accumulation of mutations, in particular C to T and A to G transversions (13,29). The

mutagenic effect is associated with loss of infectivity and with an effective clinical response (29). In vitro studies have not shown lethal mutagenesis to be a feature of remdesivir (30,31). However, to exclude non-lethal mutagenesis as a possible explanation for continuing high viral RNAs despite remdesivir treatment, we compared the mutational burden and patterns of transitions and transversions in treated and untreated patients. In accordance with current understanding of its mode of action, we found neither an increased mutational burden in remdesivir-treated patients nor any evidence of associated mutational signature (supplementary figure 5 and supplementary figure 6). We found no evidence that the proportion of transitions and transversions distribution of mutation was dependent on the treatment (Fisher t-test,  $p=0.13$ ).

#### *Measurement of subgenomic RNA*

To determine whether, despite stable viral ct values, remdesivir treatment may have inhibited viral replication we analysed subgenomic RNA (sgRNA) detected by the genome sequencing using the program Periscope (32). sgRNA derives from negative stranded RNA intermediates generated during SARS-CoV-2 replication (33,34). SARS-CoV-2 elaborates ten canonical transcripts, although ORF 10 sgRNA has never been detected (32,35–37). To optimise recovery, we excluded genomes with <90% coverage and <100 mean read depth (MRD). Stacked bar-plots of the frequency of sgRNA reads per 100,000 mapped reads (sgRPHT) for each gene and corresponding ct values for each patient are compared in figure 3A. Unlike previous reports, no correlation was found between sgRPHT and ct-values (supplementary figure 7A) (35). We have excluded patient D from the comparative analysis below as their virus was suppressed below the limit of detection during remdesivir treatment, which does not permit us to measure sgRPHT by the current technique. sgRNA levels in samples taken during remdesivir treatment for patients A and G were lower than in samples off treatment (Mann-



Whitney-Wilcoxon Test  $p = 0.05$ ) (figure 3B). Samples 6,7 and 8 in patient A had a mean viral ct values of 28.7 (range 28.24 – 29.02) with barely detectable sgRNA despite high MRDs (3402-6297). sgRNA was detected in all other samples with ct values  $<35$  bar two. A similar tendency towards significance of remdesivir treatment on levels of sgRPHT was shown across all patients over all time points (Mann-Whitney-Wilcoxon Test  $p = 0.059$ ) (supplementary figure 7 B).

### *Evidence of mixed infection*

We observed changes in the consensus sequences at different timepoints in patients A, H and I as depicted in figure 2A. To examine this and the identical sequences from patients D and A further, we analysed minor variant alleles (MVAs) as outlined in the methods. A complete list of identified polymorphisms for each sample, and corresponding frequency and read support, can be found in supplementary table 3. Patients A, B, H, and I had well supported MVAs which varied in frequency over time (supplementary figure 8). MVAs in other patients occurred for the most part on a single occasion or at levels  $<20\%$  with poor read support. To resolve possible mixed infections within each sample, we used the haplotype reconstruction method HaROLD (38). HaROLD identified three haplotypes for patient A, four for patient H, two for patient I, two in patient B with one in all other patients. All identified haplotypes are labelled as Hap\_[number]\_[Patient][Time]. Haplotypes clustered phylogenetically by patient other than for Hap1\_A, found at the root of the clade from patient A, which was identical to viral sequences from patient D (figure 2B). Haplotype-defining sites are indicated along branches, the corresponding read support for each of these is identified in supplementary table 3. The abundance of each haplotype over time is shown for patients A, B, H, and I (figure 4 A). We observed no obvious pattern of haplotype change in any of the patients, with sample-to-sample variation in viral loads and haplotype abundance occurring particularly in patients A and H.

Since independent SARS-CoV-2 lineages that differ by one or two SNPs are frequently seen in the UK, we first determined whether the haplotypes within patients A, H, I, and B were likely to represent co-infections with different viruses (39). Using Local Lineage and Monophyly Assessment (LLAMA) (40), we identified the lineage in the global alignment on COVID-19 Genomics UK (**COG-UK**) (41) consortium nearest to each haplotype. The local trees identified by LLAMA are shown in supplementary figure 9, with a single local tree representing each of the observed lineages. Comparison of haplotypes in patients A, H, and I with global sequences confirmed that haplotypes Hap1\_A, Hap2\_A, Hap3\_H and Hap1\_I were circulating as independent lineages in the UK and globally as were the single genotypes from patients C, E, and F (Supplementary Figure 9). However, haplotypes Hap3\_A, Hap1/2\_B, Hap1/2\_H and Hap2\_I as well as the virus from patient G were not represented among 61740 global sequences available on GISAID (42). This may reflect incomplete population sampling or the recent emergence of new lineages. Alternatively, it is possible that some haplotypes e.g. Hap3\_A, Hap1/2\_H and Hap2\_I are not currently freely circulating in the population and possibly that the mutations they carry are deleterious as has been postulated for influenza(43). The close clustering between each of the non-circulating viruses with other haplotypes in their cognate hosts supports the possibility of within-host evolution (supplementary figure 9).

An alternative explanation, given that patients A, B, H, and I apparently had mixed populations on first testing or soon after, is that these patients were co-infected with multiple viruses as has previously been suggested (44). Analysis of a separate dataset of 318 sequences collected in Glasgow between 28/02/2020 and 29/04/2020 found no relationship between the number of shared MVAs and the consensus sequence pairwise distance, suggesting that shared diversity is not commonly due to co-transmission in contrast to earlier work (supplementary figure 10)

(44). As proof of concept we reconstructed haplotypes for four of the sequences, each of which had optimal MVA distribution (as detailed in the methods) for HaROLD analysis. We found between two and three closely related haplotypes for each patient clustered by patient as described for samples A, B, H, and I (supplementary figure 10). As was the case for A, H and I, only one haplotype per patient was represented among freely circulating viruses in global sequence alignment provided by COG-UK, reinforcing the likelihood that the variant genotypes had arisen within each host.

A possible exception to this conclusion was patient A. It is conceivable that haplotype Hap1\_A, which is identical to the viral sequence from patient D, could have resulted from superinfection. Patients A and D were co-located in the Intensive Care Unit (ICU) (Supplementary Figure 1). Haplotype Hap1\_A was not identified in the sample taken from patient A before their transfer to the ICU but was detected approximately 28 hours later in a sample taken two hours before starting remdesivir (Figure 4 B). Were superinfection the explanation, the appearance of Hap2\_A could be explained either by recombination between Hap3\_A and Hap1\_D following transfer of patient A to the ICU or by within-host mutation of Hap 3\_A. However, no other healthcare associated transmissions were reported in the ICU and local epidemiological investigation made it unlikely. Moreover, in view of the rapid appearance of Hap1\_A following transfer to ICU and the variation in haplotype frequencies over time, it is perhaps more probable that Hap1\_A and 2\_A were present in the patient's first sample but not detected. This would support within host evolution of these three closely related strains, for one of which, Hap3\_A, there is no evidence of circulation in the population.

## Discussion

Our comparative analysis of longitudinal samples from remdesivir-treated and untreated patients infected with SARS-CoV-2 identifies evidence of remdesivir-associated suppression of viral RNA and sgRNA *in vivo* and uncovers the presence of mixed viral haplotypes likely to have evolved within each host early in infection, persisting thereafter, in some cases possibly within discrete tissue compartments in the lung.

Suppression of viral replication is a recognised goal of SARS-CoV-2 treatment, since better clinical outcomes have been associated with lower viral loads (14,45). While several drugs with *in vitro* activity against SARS-CoV-2 have been identified, only remdesivir has shown any clinical promise(10). However, reports of its efficacy are mixed (3,7–9,11) and only one other report of it having an effect on viral replication *in vivo* has been described (46). We show evidence of clinical improvement in all three patients whilst on remdesivir. In patient D clinical improvement was associated with a fall in viral RNA levels within two days of starting remdesivir. Viral RNA remained undetectable until treatment was stopped, when it again rose. In patient A no change in viral RNA levels occurred during remdesivir treatment, however, SARS-CoV-2 sgRNA levels appear to have been reduced, increasing again following cessation of treatment (figure 4). sgRNA levels have been shown to have a weak association with viral replication in *in vitro* culture (47), however the extent to which sgRNA levels, mainly of the E gene which are most often measured, reflect viable virus better than viral RNA remains controversial (35). Overall, for samples obtained during remdesivir therapy we observe lower sgRNA levels even in the presence of high viral RNA (figure 4). These samples have good coverage and MRDs (supplementary table 1), thus excluding RNA degradation (32). Our results mirror findings from the macaque model wherein animals are treated with remdesivir 12 hours following infection with SARS-CoV-2. In this model the viability in *in vitro* culture

of virus from upper respiratory samples is decreased and clinical scores are improved, despite no change in viral RNA levels (48). The possibility that sgRNA, may, together with viral load, provide a biomarker of response to remdesivir should now be explored.

We have previously used deep sequencing and evolutionary modelling of influenza and norovirus genomes to gain insight into the *in vivo* action and efficacy of ribavirin and favipiravir, drugs with similar modes of action to remdesivir (13,29). Using a similar approach, we show considerable variation in the consensus SARS-CoV-2 sequences obtained from four of the nine patients, which appears unrelated to remdesivir treatment. The heterogeneity is explained by the presence of between 2-4 stable viral haplotypes that vary in abundance (figures 4A). From this we see that, while remdesivir suppressed viral replication in patient D, the same dose had little impact on the viral RNA from the identical strain, Hap1\_A, in patient A (figure 4B). Remdesivir resistance was not found to account for the difference in response. Instead, the variable abundance of multiple distinct haplotypes present in different samples from patients A, B, H, and I, including in samples obtained from deep within the lung, is consistent with the presence of tissue compartmentalisation, wherein a pathogen replicates in physically separated niches within the lung potentially accumulating mutations that allow different populations to be distinguished. Diagnostic specimens that incompletely sample these poorly mixed virus populations, result in mixed haplotypes that vary in frequency from sample-to-sample, as demonstrated in our patient data. The findings are supported by the independent finding that SARS-CoV-2 viral loads vary in different areas of lung sampled post mortem (49).

Tissue compartmentalisation is well described for other inflammatory lung infections including *Mycobacterium tuberculosis* (M.tb) and influenza and has been associated with uneven drug penetration, leading to poor resolution of infection and predisposing to drug resistance (29,50,51). Modelling of drug levels within lung tissue from SARS-CoV-2 infected patients

suggests that remdesivir penetration into lung tissue is poor (52). In patient A, the persistence of different viral populations at variable frequencies including in bronchoalveolar lavage material supports the presence of viral compartmentalisation. Other patients may also have had tissue compartmentalisation but, in the absence of distinct viral populations, this could not be detected. Another possibility is that the effective remdesivir dose was too low, for example in patient G who was obese (table 1) or that duration of treatment too short, as in patients A and D in whom viral sgRNA and RNA respectively rebounded at the end of treatment (figures 1 and 3A). Clearance of remdesivir metabolites is affected by renal function (53). Neonates, such as patient D, are known to have relatively reduced renal clearance of drugs; accumulation of metabolites may therefore have contributed to viral RNA suppression in patient D (54). Poor drug penetration into infected lung tissue has clearly been demonstrated to underlie suboptimal responses to treatment of *M.tb* infections (50,51) and tissue compartmentalisation is recognised to contribute to the emergence of drug resistance in both *Mtb* and Influenza (12,50). For both *M.tb* and Influenza, combination therapies have been shown to achieve better pathogen clearance from lung and clinical outcomes than single agents (50,55).

High fidelity deep sequencing of multiple samples from the same patient together with computational methods that discard poorly supported MVAs minimised the contribution of artefact for the interpretation of within-host diversity in our sample set (17,38). Many of the persistent MVAs were present in early samples. As seen in other RNA viruses MVAs in SARS-CoV-2 may be more common in patients with primary or secondary immunodeficiency (13,29); we did not see the accumulation of mutations over time that characterises, for example, norovirus infections in the immunocompromised (supplementary figures 5 and 6). Haplotypes occurring within an individual were phylogenetically sister taxa, with generally only one haplotype per patient identical to freely circulating lineages (supplementary figure 9). Many circulating SARS-CoV-2 lineages differ by similar numbers (one to two) of SNPs as the

haplotypes found within our patients, making it possible that the latter were in fact independent strains that were too rare to have been previously sampled. Such strains would have to have been acquired through co-transmission or superinfection. Although this is possible for patient A, on balance the data do not support this. Another possibility is that mixed haplotypes represent co-transmission of different strains (56). To explore the likelihood of this scenario, we used a large, independent and well curated data set of 318 geographically co-located SARS-CoV-2 sequences. We show that sequences that were phylogenetically closer did not share more MVAs (supplementary figure 10), suggesting the co-transmission of infections is unlikely. The result was not changed by excluding strains sharing single MVAs which are the most likely to be artefact. Analysis of four randomly chosen samples with multiple well supported MVAs at levels of 15-85% (optimal for haplotype reconstruction), generated identical findings to those of our patients, namely sister taxa within hosts with only one viral sequence per patient found among freely circulating lineages. We conclude that early within-host evolution most logically explains the diversity seen in patients A, B, H, and I. Within evolution of stable viral haplotypes (clones) is well described for norovirus and influenza in immunocompromised patients (57,58) Patients B, H and I were known to be immunocompromised, while Patient A remains under investigation for immune dysfunction. In contrast none of the four adult patients with mixed haplotypes from the independent cohort, were classified as immunocompromised (data not shown). More extensive analysis to understand better the frequency with which within-host variants evolve, why they arise in some patients and not others and whether there are any correlations with underlying clinical condition or outcomes is ongoing.

A major caveat to our findings is that our cohort is paediatric, for which both the clinical picture and outcome of SARS-CoV-2 infections are known to differ from adults. Notwithstanding, similar patterns of clinical and virological response to remdesivir have been described in adults

(3,10,46,59). Moreover, both clinical and viral sequence data from the use of repurposed drugs to treat other severe RNA viral infections has shown similarities in adults and children (12,50,55). Larger studies using deep clinical and viral profiling of multiple samples from adult patients treated with remdesivir alone and in combination would provide better insight.

In summary, we show that treatment with remdesivir is capable of reducing SARS-CoV-2 viral and subgenomic RNA *in vivo* and demonstrate that the latter, in particular, needs further investigation as a potential biomarker for monitoring antiviral therapy. Our data suggest that heterogeneous response to remdesivir is likely in many cases to be due to non-viral factors including, potentially, inadequate dosing and duration of treatment. The patterns of SARS-CoV-2 within-host genetic heterogeneity uncovered by deep sequencing may be most parsimoniously explained by viral compartmentalisation within lung-tissue, a factor that is already known to impede drug penetration in patients with other lung infections. This may compound inherently poor remdesivir tissue penetration and rapid clearance of active metabolites in those with normal renal function. We and others have shown that where compartmentalisation occurs in influenza (29) and *M. tuberculosis* (50), combination therapies improve outcomes. Based on our experience of using similar drugs for the treatment of serious RNA viral infections, we propose that a more personalised medicine approach combining *in vitro* and *in vivo* pharmacokinetic measurements, viral RNA and sgRNA profiling together with viral evolutionary modelling is likely to yield important insights into the optimal use of remdesivir both alone and in combination for treatment of SARS-CoV-2. Finally, our data point to the presence of mixed genotypes in a subset of patients, which are most likely explained by within-host evolution rather than co-transmission.



## Materials and Methods

### *Study Cohort and approval*

We sequenced SARS-CoV-2 samples routinely collected for clinical monitoring from children hospitalized for COVID-19 between early March and mid-May 2020. This study was approved by Great Ormond Street Hospital (Clinical Audit Number #2857) and PHE Research Ethics and Governance Group (REGG) (R&D NR0195).

### *Sample Collection and Viral Sequencing*

Nasopharyngeal swab samples were collected and tested for SARS-CoV-2. Full length SARS-CoV-2 genome sequences were obtained from all positive samples using SureSelect<sup>XT</sup> target enrichment and Illumina sequencing. Reads generated were trimmed with trim galore (60). Duplicated reads were removed using Picard (61). For each patient, a unique patient reference was generated by mapping the remaining reads of the first sample to the SARS-CoV-2 reference genome (NC\_045512) from GenBank using bwa-mem (62). The mapping quality was checked using Qualimap and the consensus whole genome sequence was generated using QUASR (63,64). Reads from the subsequent samples of the same patient were mapped to this patient reference. Consensus sequences were aligned using MAFFT (65). Only genomes with more than 80 percent genome coverage and a mean read depth of 100 or above were included in downstream analysis. Seqtk was used to subsample the reads for normalisation (66).

### *Phylogenetic Analysis*

Maximum Likelihood tree of the alignment was constructed using RAxML (67), with the GTR model and 1000 bootstrap replicates. All trees were rooted on the SARS-CoV-2 reference genome NC\_045512.

### *Analysis and figure generation*

Analysis was completed in R 3.6.1 using Rstudio 1.2 (68,69). In general data was processed using dplyr (v0.8.3), figures were generated using ggplot2 (v3.3.1), both part of the tidyverse family of packages (v1.2.1) (70). We employed the fisher.test in R to compare the count-data of mutations for treated and untreated samples across all individuals(68). The Mann-Whitney-Wilcoxon Test was implemented using the wilcox.test in R (68). We used Pearson's and Spearman's rank correlation for correlation analysis. This was done using the lm.test and cor.test function in the stats package in R (68).

### *Haplotype Reconstruction*

Haplotypes were reconstructed using HAplotype Reconstruction Of Longitudinal Deep Sequences (HaROLD) with default settings (38). HaROLD does not statistically support haplotypes where from a single MVA, we therefore constructed by hand the haplotypes for Patient B. In this case the haplotype frequency was taken to be that of the single MVA.

### *Quantification of subgenomic RNA*

We employed Periscope to detect subgenomic RNA (sgRNA) (32). sgRNA is identified based on the detection of the leader sequence at the 5'end (5'-AACCAACTTTCGATCTCTTG TAGATCTGTTCT-3') of the sequence.

### *Minority Variant Calling*

Minority allele variants had to have a frequency of above 2% and with a minimum of 4 supporting reads identified at sites with a read depth of  $\geq 5$  using VarScan (71). Transient MAVs, which occurred at one timepoint in an individual were discarded from the analysis.

### *Glasgow COG-UK Sequencing and Sample Selection for haplotype reconstruction*

Minority variants as described above, were called for 318 sequences from Glasgow in the COG-UK from samples collected between 28/02/2020 and 29/04/2020. Following extraction, samples were DNase treated (AM2222) and the libraries were prepared utilising the protocols developed by the ARTIC network (<https://artic.network/ncov-2019>). Amplicons were generated using either primer sets v1, v2 and v3. The DNA amplicon fragments were cleaned using AMPURE beads (Beckman Coulter) and 40 ng was used to prepare the Illumina sequencing libraries with a DNA KAPA library preparation kit (Roche). Indexing was carried out with NEBNext multiplex oligos (NEB) with 7 cycles of PCR. Libraries were pooled in equimolar amounts and loaded on a MiSeqV2 cartridge (500 cycles). In the case of CR111, the library was sequenced on Illumina's NextSeq 550 System (Illumina, Part Number SY-415-1002).

To optimise the performance of HaROLD we selected samples with mean read depths of >500 and SNVs occurring at frequencies of >15% with both alleles supported by >100 reads. The number of MVAs was tallied for all samples, and those that reported numbers in the top 2% of all samples were carried forward for analysis by Harold.

### *Structural Biology*

The structure of the spike protein PDB 6XR8 was visualised using VMD (72,73). Mutations were modelled using Swissmodel (74,75).

## Supplementary Materials

### Supplementary Tables

1. Supplementary Table 1: Sequencing Metrics.
2. Supplementary Table 2: Summary of consensus level mutations in patients A, H, and I.
3. Summary of variants detected in all patients.

### Supplementary Figures

1. Supplementary Figure 1: Patient timeline overview
2. Supplementary Figure 2: Ct trajectories of 9 patients for all samples collected.
3. Supplementary Figure 3: Clinical markers for 9 patients.
4. Supplementary Figure 4: The P812L mutation does not affect the structure of the SARs-CoV-2 spike protein.
5. Supplementary Figure 5: Mutagenic signature for each patient over time.
6. Supplementary Figure 6: Mutational burden over time in each patient.
7. Supplementary Figure 7: Comparison of sgRNA reads per 100,000 mapped reads (sgRPHT) in treated and untreated individuals.
8. Supplementary Figure 8: Polymorphisms trajectories for each patient.
9. Supplementary Figure 9: Comparison of haplotypes from patient A , B, H and I and consensus of C, D, E, G with global sequences.
10. Supplementary Figure 10: Scatter plot of shared polymorphisms (or MVAs) vs pairwise distance for 318 from Glasgow collected between 28/02/2020 and 29/04/2020.
11. Supplementary figure 11: Frequency of identified haplotypes from 4/6 Glasgow samples processed through Harold.

## References

1. World Health Organization. WHO Timeline - COVID-19 [Internet]. [cited 2020 May 30]. Available from: <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>
2. WHO Coronavirus Disease (COVID-19) Dashboard [Internet]. [cited 2020 Sep 28]. Available from: <https://covid19.who.int>
3. Y Wang, D Zhang, G Du, R Du, J Zhao, Y Jin, S Fu, L Gao, Z Cheng, Q Lu, Y Hu, G Luo, K Wang, Y Lu, H Li, S Wang, S Ruan, C Yang, C Mei, Y Wang, D Ding, F Wu, X Tang, X Ye, Y Ye, B Liu, J Yang, W Yin, A Wang, G Fan, F Zhou, Z Liu, X Gu, J Xu, L Shang, Y Zhang, L Cao, T Guo, Y Wan, H Qin, Y Jiang, T Jaki, FG Hayden, PW Horby, B Cao, C Wang. Remdesivir in adults with severe COVID-19: a randomised, double-blind, placebo-controlled, multicentre trial. *The Lancet*. 2020 May;395(10236):1569–78.
4. KK-W To, OT-Y Tsang, W-S Leung, AR Tam, T-C Wu, DC Lung, CC-Y Yip, J-P Cai, JM-C Chan, TS-H Chik, DP-L Lau, CY-C Choi, L-L Chen, W-M Chan, K-H Chan, JD Ip, AC-K Ng, RW-S Poon, C-T Luo, VC-C Cheng, JF-W Chan, IF-N Hung, Z Chen, H Chen, K-Y Yuen. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *The Lancet Infectious Diseases*. 2020 Mar;S1473309920301961.
5. F Yu, L Yan, N Wang, S Yang, L Wang, Y Tang, G Gao, S Wang, C Ma, R Xie, F Wang, C Tan, L Zhu, Y Guo, F Zhang. Quantitative Detection and Viral Load Analysis of SARS-CoV-2 in Infected Patients. *Clinical Infectious Diseases*. 2020 Jul 28;71(15):793–8.
6. Y Pan, D Zhang, P Yang, LLM Poon, Q Wang. Viral load of SARS-CoV-2 in clinical samples. *The Lancet Infectious Diseases*. 2020 Apr;20(4):411–2.

7. J Grein, N Ohmagari, D Shin, G Diaz, E Asperges, A Castagna, T Feldt, G Green, ML Green, F-X Lescure, E Nicastrì, R Oda, K Yo, E Quiros-Roldan, A Studemeister, J Redinski, S Ahmed, J Bennett, D Chelliah, D Chen, S Chihara, SH Cohen, J Cunningham, A D'Arminio Monforte, S Ismail, H Kato, G Lapadula, E L'Her, T Maeno, S Majumder, M Massari, M Mora-Rillo, Y Mutoh, D Nguyen, E Verweij, A Zoufaly, AO Osinusi, A DeZure, Y Zhao, L Zhong, A Chokkalingam, E Elboudwarej, L Telep, L Timbs, I Henne, S Sellers, H Cao, SK Tan, L Winterbourne, P Desai, R Mera, A Gaggar, RP Myers, DM Brainard, R Childs, T Flanigan. Compassionate Use of Remdesivir for Patients with Severe Covid-19. *New England Journal of Medicine*. 2020 Jun 11;382(24):2327–36.
8. SA Olender, KK Perez, AS Go, B Balani, EG Price-Haywood, NS Shah, S Wang, TL Walunas, S Swaminathan, J Slim, B Chin, S De Wit, SM Ali, A Soriano Viladomiu, P Robinson, RL Gottlieb, TYO Tsang, I-H Lee, H Hu, RH Haubrich, AP Chokkalingam, L Lin, L Zhong, BN Bekele, R Mera-Giler, C Phulpin, H Edgar, J Gallant, H Diaz-Cuervo, LE Smith, AO Osinusi, DM Brainard, JI Bernardino, GS-US-540–5773 and GS-US-540–5807 Investigators. Remdesivir for Severe Coronavirus Disease 2019 (COVID-19) Versus a Cohort Receiving Standard of Care. *Clinical Infectious Diseases*. 2020 Jul 24;ciaa1041.
9. JD Norrie. Remdesivir for COVID-19: challenges of underpowered studies. *The Lancet*. 2020 May;395(10236):1525–7.
10. JH Beigel, KM Tomashek, LE Dodd, AK Mehta, BS Zingman, AC Kalil, E Hohmann, HY Chu, A Luetkemeyer, S Kline, D Lopez de Castilla, RW Finberg, K Dierberg, V Tapson, L Hsieh, TF Patterson, R Paredes, DA Sweeney, WR Short, G Touloumi, DC Lye, N Ohmagari, M Oh, GM Ruiz-Palacios, T Benfield, G Fätkenheuer, MG Kortepeter, RL Atmar, CB Creech, J Lundgren, AG Babiker, S Pett, JD Neaton, TH Burgess, T

- Bonnett, M Green, M Makowski, A Osinusi, S Nayak, HC Lane. Remdesivir for the Treatment of Covid-19 — Final Report. *New England Journal of Medicine*. 2020 May 22;0(0):null.
11. WS trial Consortium, H Pan, R Peto, QA Karim, M Alejandria, AM Henao-Restrepo, CH García, M-P Kieny, R Malekzadeh, S Murthy, M-P Preziosi, S Reddy, MR Periago, V Sathiyamoorthy, J-A Røttingen, S Swaminathan, AR for the C and I of TA as the members of the Writing Committee. Repurposed antiviral drugs for COVID-19 –interim WHO SOLIDARITY trial results. *medRxiv*. 2020 Oct 15;2020.10.15.20209817.
  12. CK Lumby, L Zhao, M Oporto, T Best, H Tutill, D Shah, P Veys, R Williams, A Worth, CJR Illingworth, J Breuer. Favipiravir and Zanamivir Cleared Infection with Influenza B in a Severely Immunocompromised Child. *Clin Infect Dis [Internet]*. [cited 2020 Sep 28]; Available from: <https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa023/5775454>
  13. C Ruis, L-AK Brown, S Roy, C Atkinson, R Williams, SO Burns, E Yara-Romero, M Jacobs, R Goldstein, J Breuer, DM Lowe. Mutagenesis in Norovirus in Response to Favipiravir Treatment. *New England Journal of Medicine*. 2018 Nov 29;379(22):2173–6.
  14. Y Liu, L-M Yan, L Wan, T-X Xiang, A Le, J-M Liu, M Peiris, LLM Poon, W Zhang. Viral dynamics in mild and severe cases of COVID-19. *The Lancet Infectious Diseases*. 2020 Jun;20(6):656–7.
  15. Y Peng, AJ Mentzer, G Liu, X Yao, Z Yin, D Dong, W Dejnirattisai, T Rostron, P Supasa, C Liu, C López-Camacho, J Slon-Campos, Y Zhao, DI Stuart, GC Paesen, JM Grimes, AA Antson, OW Bayfield, DEDP Hawkins, D-S Ker, B Wang, L Turtle, K Subramaniam,

- P Thomson, P Zhang, C Dold, J Ratcliff, P Simmonds, T de Silva, P Sopp, D Wellington, U Rajapaksa, Y-L Chen, M Salio, G Napolitani, W Paes, P Borrow, BM Kessler, JW Fry, NF Schwabe, MG Semple, JK Baillie, SC Moore, PJM Openshaw, MA Ansari, S Dunachie, E Barnes, J Frater, G Kerr, P Goulder, T Lockett, R Levin, Y Zhang, R Jing, L-P Ho, RJ Cornall, CP Conlon, P Klenerman, GR Screaton, J Mongkolsapaya, A McMichael, JC Knight, G Ogg, T Dong. Broad and strong memory CD4 + and CD8 + T cells induced by SARS-CoV-2 in UK convalescent individuals following COVID-19. *Nature Immunology*. 2020 Nov;21(11):1336–45.
16. L van Dorp, M Acman, D Richard, LP Shaw, CE Ford, L Ormond, CJ Owen, J Pang, CCS Tan, FAT Boshier, AT Ortiz, F Balloux. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infection, Genetics and Evolution*. 2020 Sep 1;83:104351.
  17. N De Maio, Walker, Conor, Borges, Rui, Weilguny, Lukas, Slodkowitz, Greg, Goldman, Nick. Issues with SARS-CoV-2 sequencing data.
  18. ML Agostini, EL Andres, AC Sims, RL Graham, TP Sheahan, X Lu, EC Smith, JB Case, JY Feng, R Jordan, AS Ray, T Cihlar, D Siegel, RL Mackman, MO Clarke, RS Baric, MR Denison. Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. Subbarao K, editor. *mBio*. 2018 Mar 6;9(2):e00221-18, /mbio/9/2/mBio.00221-18.atom.
  19. CM Poh, G Carissimo, B Wang, SN Amrun, CY-P Lee, RS-L Chee, S-W Fong, NK-W Yeo, W-H Lee, A Torres-Ruesta, Y-S Leo, MI-C Chen, S-Y Tan, LYA Chai, S Kalimuddin, SSG Kheng, S-Y Thien, BE Young, DC Lye, BJ Hanson, C-I Wang, L Renia, LFP Ng. Two linear epitopes on the SARS-CoV-2 spike protein that elicit



- neutralising antibodies in COVID-19 patients. *Nature Communications*. 2020 Jun 1;11(1):2806.
20. A Grifoni, J Sidney, Y Zhang, RH Scheuermann, B Peters, A Sette. A Sequence Homology and Bioinformatic Approach Can Predict Candidate Targets for Immune Responses to SARS-CoV-2. *Cell Host & Microbe*. 2020 Apr 8;27(4):671-680.e2.
  21. M Zheng, L Song. Novel antibody epitopes dominate the antigenicity of spike glycoprotein in SARS-CoV-2 compared to SARS-CoV. *Cellular & Molecular Immunology*. 2020 May;17(5):536–8.
  22. SF Ahmed, AA Quadeer, MR McKay. Preliminary Identification of Potential Vaccine Targets for the COVID-19 Coronavirus (SARS-CoV-2) Based on SARS-CoV Immunological Studies. *Viruses*. 2020 Mar;12(3):254.
  23. C Wu, Y Liu, Y Yang, P Zhang, W Zhong, Y Wang, Q Wang, Y Xu, M Li, X Li, M Zheng, L Chen, H Li. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharmaceutica Sinica B*. 2020 May 1;10(5):766–88.
  24. CJ Gordon, EP Tchesnokov, JY Feng, DP Porter, M Götte. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J Biol Chem*. 2020 Apr 10;295(15):4773–9.
  25. M Macchiagodena, M Pagliai, P Procacci. Identification of potential binders of the main protease 3CLpro of the COVID-19 via structure-based ligand design and molecular modeling. *Chemical Physics Letters*. 2020 Jul 1;750:137489.

26. S Durdagi, B Aksoydan, B Dogan, K Sahin, A Shahraki. Screening of Clinically Approved and Investigation Drugs as Potential Inhibitors of COVID-19 Main Protease: A Virtual Drug Repurposing Study. 2020 Mar 26 [cited 2020 Sep 28]; Available from: [/articles/preprint/Screening\\_of\\_Clinically\\_Approved\\_and\\_Investigation\\_Drugs\\_as\\_Potential\\_Inhibitors\\_of\\_COVID-19\\_Main\\_Protease\\_A\\_Virtual\\_Drug\\_Repurposing\\_Study/12032712/1](#)
27. KG Andersen, A Rambaut, WI Lipkin, EC Holmes, RF Garry. The proximal origin of SARS-CoV-2. *Nature Medicine*. 2020 Apr;26(4):450–2.
28. B Coutard, C Valle, X de Lamballerie, B Canard, NG Seidah, E Decroly. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral Research*. 2020 Apr 1;176:104742.
29. CK Lumby, L Zhao, M Oporto, T Best, H Tutill, D Shah, P Veys, R Williams, A Worth, CJR Illingworth, J Breuer. Favipiravir and Zanamivir Cleared Infection with Influenza B in a Severely Immunocompromised Child. *Clinical Infectious Diseases*. 2020 Mar 3;ciaa023.
30. SJF Kaptein, S Jacobs, L Langendries, L Seldeslachts, S Ter Horst, L Liesenborghs, B Hens, V Vergote, E Heylen, K Barthelemy, E Maas, C De Keyzer, L Bervoets, J Rymenants, T Van Buyten, X Zhang, R Abdelnabi, J Pang, R Williams, HJ Thibaut, K Dallmeier, R Boudewijns, J Wouters, P Augustijns, N Verougstraete, C Cawthorne, J Breuer, C Solas, B Weynand, P Annaert, I Spriet, G Vande Velde, J Neyts, J Rocha-Pereira, L Delang. Favipiravir at high doses has potent antiviral activity in SARS-CoV-2-infected hamsters, whereas hydroxychloroquine lacks activity. *Proc Natl Acad Sci U S A*. 2020 Oct 9;

31. EP Tchesnokov, JY Feng, DP Porter, M Götte. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. *Viruses*. 2019 04;11(4).
32. MD Parker, BB Lindsey, S Leary, S Gaudieri, A Chopra, M Wyles, A Angyal, LR Green, P Parsons, RM Tucker, R Brown, D Groves, K Johnson, L Carrilero, J Heffer, D Partridge, C Evans, M Raza, AJ Keeley, N Smith, D Wang, S Mallal, TI de Silva. periscope: sub-genomic RNA identification in SARS-CoV-2 ARTIC Network Nanopore Sequencing Data [Internet]. *Bioinformatics*; 2020 Jul [cited 2020 Oct 1]. Available from: <http://biorxiv.org/lookup/doi/10.1101/2020.07.01.181867>
33. I Sola, F Almazán, S Zúñiga, L Enjuanes. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol*. 2015 Nov;2(1):265–88.
34. DF Stern, SIT Kennedy. Coronavirus Multiplication Strategy I. Identification and Characterization of Virus-Specified RNA. *Journal of Virology*. 1980 Jun 1;34(3):665–74.
35. S Alexandersen, A Chamings, TR Bhatta. SARS-CoV-2 genomic and subgenomic RNAs in diagnostic samples are not an indicator of active replication. *medRxiv*. 2020 Aug 16;2020.06.01.20119750.
36. G Taiaroa, D Rawlinson, L Featherstone, M Pitt, L Caly, J Druce, D Purcell, L Harty, T Tran, J Roberts, N Scott, M Catton, D Williamson, L Coin, S Duchene. Direct RNA sequencing and early evolution of SARS-CoV-2. *bioRxiv*. 2020 Apr 3;2020.03.05.976167.
37. The Architecture of SARS-CoV-2 Transcriptome: Cell [Internet]. [cited 2020 Oct 12]. Available from: [https://www.cell.com/cell/fulltext/S0092-8674\(20\)30406-2?\\_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867420304062%3Fshowall%3Dtrue](https://www.cell.com/cell/fulltext/S0092-8674(20)30406-2?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867420304062%3Fshowall%3Dtrue)

38. J Pang, C Venturini, AU Tamuri, S Roy, J Breuer, RA Goldstein. Haplotype assignment of longitudinal viral deep-sequencing data using co-variation of variant frequencies [Internet]. Genomics; 2020 Aug [cited 2020 Sep 3]. Available from: <http://biorxiv.org/lookup/doi/10.1101/444877>
39. A Rambaut, EC Holmes, Á O'Toole, V Hill, JT McCrone, C Ruis, L du Plessis, OG Pybus. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*. 2020 Nov;5(11):1403–7.
40. O'Toole, Aine. Local Lineage and Monophyly Assessment [Internet]. Available from: <https://github.com/cov-lineages/llama>
41. An integrated national scale SARS-CoV-2 genomic surveillance network. *The Lancet Microbe*. 2020 Jul;1(3):e99–100.
42. S Elbe, G Buckland-Merrett. Data, disease and diplomacy: GISAID's innovative contribution to global health: Data, Disease and Diplomacy. *Global Challenges*. 2017 Jan;1(1):33–46.
43. KS Xue, JD Bloom. Linking influenza virus evolution within and between human hosts. *Virus Evol* [Internet]. 2020 Jan 1 [cited 2020 Sep 28];6(1). Available from: <https://academic.oup.com/ve/article/6/1/veaa010/5739536>
44. KA Lythgoe, M Hall, L Ferretti, M de Cesare, G MacIntyre-Cockett, A Trebes, M Andersson, N Otecko, EL Wise, N Moore, J Lynch, S Kidd, N Cortes, M Mori, A Justice, A Green, MA Ansari, L Abeler-Dörner, CE Moore, TEA Peto, R Shaw, P Simmonds, D Buck, JA Todd, on behalf of OA Group, D Bonsall, C Fraser, T Golubchik. Shared SARS-CoV-2 diversity suggests localised transmission of minority variants. *bioRxiv*. 2020 May 29;2020.05.28.118992.

45. JH Beigel, KM Tomashek, LE Dodd, AK Mehta, BS Zingman, AC Kalil, E Hohmann, HY Chu, A Luetkemeyer, S Kline, D Lopez de Castilla, RW Finberg, K Dierberg, V Tapson, L Hsieh, TF Patterson, R Paredes, DA Sweeney, WR Short, G Touloumi, DC Lye, N Ohmagari, M Oh, GM Ruiz-Palacios, T Benfield, G Fätkenheuer, MG Kortepeter, RL Atmar, CB Creech, J Lundgren, AG Babiker, S Pett, JD Neaton, TH Burgess, T Bonnett, M Green, M Makowski, A Osinusi, S Nayak, HC Lane. Remdesivir for the Treatment of Covid-19 — Preliminary Report. *New England Journal of Medicine*. 2020 May 22;0(0):null.
46. MS Buckland, J Galloway, CN Fhogartaig, L Meredith, NM Provine, S Bloor, A Ogbe, WM Zelek, A Smielewska, A Yakovleva, T Mann, L Bergamaschi, L Turner, F Mescia, EJM Toonen, C-P Hackstein, HD Akther, VA Vieira, L Ceron-Gutierrez, J Periselneris, S Kiani-Alikhan, S Grigoriadou, D Vaghela, SE Lear, E Torok, WL Hamilton, J Quick, J Stockton, P Nelson, M Hunter, TI Coulter, L Devlin, J Bradley, K Smith, W Ouwehand, L Estcourt, HH Simmonds, D Roberts, I Wilkinson, N Sreaton, N Loman, P Lyons, R Doffinger, P Morgan, I Goodfellow, P Klenerman, P Lehner, N Matheson, J Thaventhiran. Successful treatment of COVID-19 with remdesivir in the absence of humoral immunity [Internet]. In Review; 2020 Sep [cited 2020 Oct 26]. Available from: <https://www.researchsquare.com/article/rs-79022/v1>
47. RAPM Perera, E Tso, OTY Tsang, DNC Tsang, K Fung, YWY Leung, AWH Chin, DKW Chu, SMS Cheng, LLM Poon, VWM Chuang, M Peiris. Early Release - SARS-CoV-2 Virus Culture and Subgenomic RNA for Respiratory Specimens from Patients with Mild Coronavirus Disease - Volume 26, Number 11—November 2020 - *Emerging Infectious Diseases* journal - CDC. [cited 2020 Sep 30]; Available from: [https://wwwnc.cdc.gov/eid/article/26/11/20-3219\\_article](https://wwwnc.cdc.gov/eid/article/26/11/20-3219_article)

48. BN Williamson, F Feldmann, B Schwarz, K Meade-White, DP Porter, J Schulz, N van Doremalen, I Leighton, CK Yinda, L Pérez-Pérez, A Okumura, J Lovaglio, PW Hanley, G Saturday, CM Bosio, S Anzick, K Barbian, T Cihlar, C Martens, DP Scott, VJ Munster, E de Wit. Clinical benefit of remdesivir in rhesus macaques infected with SARS-CoV-2. *Nature*. 2020 Sep;585(7824):273–6.
49. N Desai, A Neyaz, A Szabolcs, AR Shih, JH Chen, V Thapar, LT Nieman, A Solovyov, A Mehta, DJ Lieb, AS Kulkarni, C Jaicks, CJ Pinto, D Juric, I Chebib, RB Colvin, AY Kim, R Monroe, SE Warren, P Danaher, JW Reeves, J Gong, EH Rueckert, BD Greenbaum, N Hachohen, SM Lagana, MN Rivera, LM Sholl, JR Stone, DT Ting, V Deshpande. Temporal and Spatial Heterogeneity of Host Response to SARS-CoV-2 Pulmonary Infection. *medRxiv* [Internet]. 2020 Aug 2 [cited 2020 Sep 28]; Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7402055/>
50. N Strydom, SV Gupta, WS Fox, LE Via, H Bang, M Lee, S Eum, T Shim, CE Barry, M Zimmerman, V Dartois, RM Savic. Tuberculosis drugs' distribution and emergence of resistance in patient's lung lesions: A mechanistic model and tool for regimen and dose optimization. Murray M, editor. *PLoS Med*. 2019 Apr 2;16(4):e1002773.
51. X Ks, M Lh, B T, B Jd. Within-Host Evolution of Human Influenza Virus [Internet]. Vol. 26, Trends in microbiology. *Trends Microbiol*; 2018 [cited 2020 Sep 28]. Available from: <https://pubmed.ncbi.nlm.nih.gov/29534854/>
52. Y Wang, L Chen. Tissue distributions of antiviral drugs affect their capabilities of reducing viral loads in COVID-19 treatment. *European Journal of Pharmacology*. 2020 Dec 15;889:173634.

53. Remdesivir in Patients with Acute or Chronic Kidney Disease and COVID-19 | American Society of Nephrology [Internet]. [cited 2020 Oct 20]. Available from: <https://jasn.asnjournals.org/content/31/7/1384>
54. E Germovsek, CIS Barker, M Sharland, JF Standing. Pharmacokinetic–Pharmacodynamic Modeling in Pediatric Drug Development, and the Importance of Standardized Scaling of Clearance. *Clin Pharmacokinet.* 2019;58(1):39–52.
55. Y Wang, G Fan, A Salam, P Horby, FG Hayden, C Chen, J Pan, J Zheng, B Lu, L Guo, C Wang, B Cao. Comparative Effectiveness of Combined Favipiravir and Oseltamivir Therapy Versus Oseltamivir Monotherapy in Critically Ill Patients With Influenza Virus Infection. *J Infect Dis.* 2020 Apr 27;221(10):1688–98.
56. Shared SARS-CoV-2 diversity suggests localised transmission of minority variants | bioRxiv [Internet]. [cited 2020 Oct 12]. Available from: <https://www.biorxiv.org/content/10.1101/2020.05.28.118992v2>
57. KS Xue, T Stevens-Ayers, AP Campbell, JA Englund, SA Pergam, M Boeckh, JD Bloom. Parallel evolution of influenza across multiple spatiotemporal scales. Neher RA, editor. *eLife.* 2017 Jun 27;6:e26875.
58. E Vega, E Donaldson, J Huynh, L Barclay, B Lopman, R Baric, LF Chen, J Vinjé. RNA Populations in Immunocompromised Patients as Reservoirs for Novel Norovirus Variants. *Journal of Virology.* 2014 Dec 15;88(24):14184–96.
59. S Gastine, J Pang, FAT Boshier, SJ Carter, DO Lonsdale, M Cortina-Borja, IFN Hung, J Breuer, F Kloprogge, JF Standing. Systematic review and patient-level meta-analysis of SARS-CoV-2 viral dynamics to model response to antiviral therapies. *medRxiv.* 2020 Aug 23;2020.08.20.20178699.

60. The Babraham Institute. TrimGalore [Internet]. 2019. Available from: <https://github.com/FelixKrueger/TrimGalore>
61. Broad Institute. Picard GitHub Repository [Internet]. 2019. Available from: <http://broadinstitute.github.io/picard/>
62. H Li, R Durbin. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009 Jul 15;25(14):1754–60.
63. K Okonechnikov, A Conesa, F García-Alcalde. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2015 Oct 1;btv566.
64. SJ Watson, MRA Welkers, DP Depledge, E Coulter, JM Breuer, MD de Jong, P Kellam. Viral population analysis and minority-variant detection using short read next-generation sequencing. *Phil Trans R Soc B*. 2013 Mar 19;368(1614):20120205.
65. K Katoh, DM Standley. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013 Apr;30(4):772–80.
66. W Shen, S Le, Y Li, F Hu. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. *PLOS ONE*. 2016 Oct 5;11(10):e0163962.
67. A Stamatakis. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014 May 1;30(9):1312–3.
68. R Core Team. R: A language and environment for statistical computing. [Internet]. Vienna, Austria.: R Foundation for Statistical Computing; 2019 [cited 2019 Sep 1]. Available from: <https://www.R-project.org/>



69. JS Racine. RStudio: A Platform-Independent IDE for R and Sweave. *Journal of Applied Econometrics*. 2012;27(1):167–72.
70. H Wickham, M Averick, J Bryan, W Chang, LD McGowan, R François, G Grolemond, A Hayes, L Henry, J Hester, M Kuhn, TL Pedersen, E Miller, SM Bache, K Müller, J Ooms, D Robinson, DP Seidel, V Spinu, K Takahashi, D Vaughan, C Wilke, K Woo, H Yutani. Welcome to the Tidyverse. *Journal of Open Source Software*. 2019 Nov 21;4(43):1686.
71. DC Koboldt, Q Zhang, DE Larson, D Shen, MD McLellan, L Lin, CA Miller, ER Mardis, L Ding, RK Wilson. VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Research*. 2012 Mar 1;22(3):568–76.
72. W Humphrey, A Dalke, K Schulten. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996 Feb 1;14(1):33–8.
73. Y Cai, J Zhang, T Xiao, H Peng, SM Sterling, RM Walsh, S Rawson, S Rits-Volloch, B Chen. Distinct conformational states of SARS-CoV-2 spike protein. *Science*. 2020 Sep 25;369(6511):1586–92.
74. S Bienert, A Waterhouse, TAP de Beer, G Tauriello, G Studer, L Bordoli, T Schwede. The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res*. 2017 04;45(D1):D313–9.
75. A Waterhouse, M Bertoni, S Bienert, G Studer, G Tauriello, R Gumienny, FT Heer, TAP de Beer, C Rempfer, L Bordoli, R Lepore, T Schwede. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res*. 2018 02;46(W1):W296–303.

76. N Alders, J Penner, K Grant, C Patterson, J Hassell, N MacDermott, S Pincott, A Bamford, P du Pré, M Johnson, K Moshal. COVID-19 Pandemic Preparedness in a United Kingdom Tertiary and Quaternary Children's Hospital: Tales of the Unexpected. *J Pediatric Infect Dis Soc* [Internet]. [cited 2020 Nov 5]; Available from: <https://academic.oup.com/jpids/advance-article/doi/10.1093/jpids/piaa135/5943885>
77. DJ Stewart, JC Hartley, M Johnson, SD Marks, P du Pré, J Stojanovic. Renal dysfunction in hospitalised children with COVID-19. *The Lancet Child & Adolescent Health*. 2020 Aug 1;4(8):e28–9.
78. RW Issitt, J Booth, WA Bryant, A Spiridou, AM Taylor, P du Pré, P Ramnarayan, J Hartley, M Cortina-Borja, K Moshal, H Dunn, H Hemingway, NJ Sebire. Children with COVID-19 at a specialist centre: initial experience and outcome. *The Lancet Child & Adolescent Health*. 2020 Aug 1;4(8):e30–1.
79. C Frauenfelder, J Brierley, E Whittaker, G Perucca, A Bamford. Infant With SARS-CoV-2 Infection Causing Severe Lung Disease Treated With Remdesivir. *Pediatrics* [Internet]. 2020 Jun 1 [cited 2020 Nov 5]; Available from: <https://pediatrics.aappublications.org/content/early/2020/06/16/peds.2020-1701>

## Acknowledgements

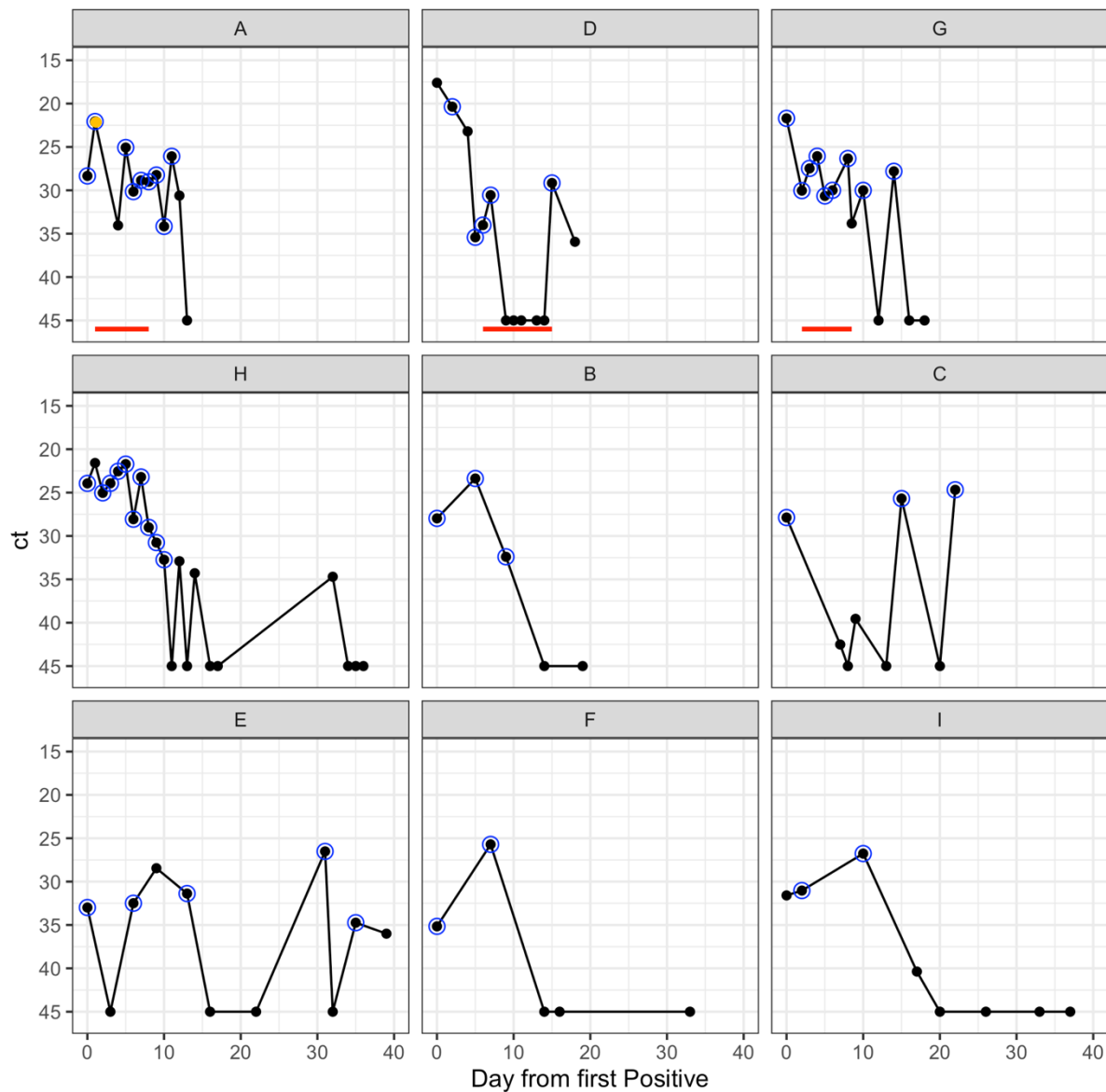
We thank the UK (COG-UK) consortium\*. COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. This work was supported by The John Black Charitable Foundation and the UCLH and GOSH NIHR biomedical research centres. JP is supported by the Rosetrees Foundation and FTB by a Wellcome Trust Collaborative Award to JB. We thank Gilead for supplying Remdesvir on the compassionate access programme and the GOSH bioethics committees together with St Mary's ID clinicians who contributed to multiplinary discussions relating to compassionate use of RDV. We would like to thank the various medical teams that cared for the patients, with special mention to: Collin Wallis, Rukshana Shroff, Philip Ancliff, Catriona Duncan, Jane Hassel, Guiseppe Barone, Mae Johnson, and Mette Jorgensen. We should also like to acknowledge the contribution of the UCL Pathogen Genomics Unit, UCL Genomics, and the Great Ormond Street Hospital Departments of Infectious Disease, and Microbiology, Virology and Infection Control and Paediatric Intensive Care. This work would not have been possible without the input from patients and their families.

\*A full list of consortium names and affiliations can be found at <https://www.cogconsortium.uk>

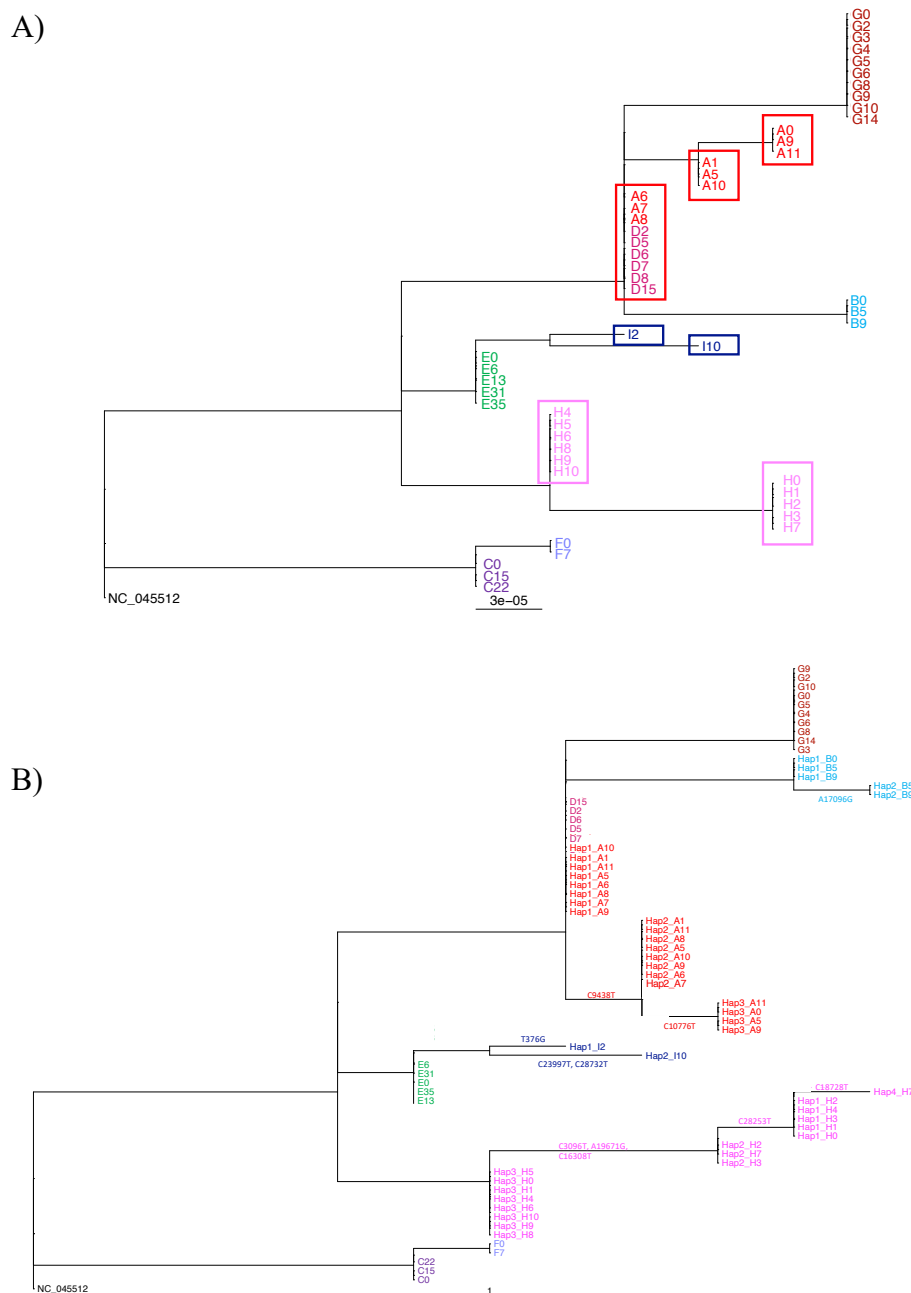
**Table 1:** Overview of patient clinical, treatment, sampling and viral lineage data for all 9 patients.

Patient	Gender	Age Group (Years)	Ethnicity	Weight (kg)	Diagnosis	Immune Status	Health Care Associated	Days from first recorded symptom to first positive	Remdesivir (Y/N) [days]	Days in ICU	No. of Sequenced Samples	Lineage	Previous Studies
A	Female	6-10	Other	39	bronchiectasis with prior lobectomy	Immunocompetent		3	Y [8]	12	9	B.1.1	(76–78)
B	Female	6-10	Asian	25.7	severe chronic illness including kidney disease	Immunocompromised		Unknown	N	0	3	B.1.1	(76–78)
C	Male	6-10	Black African	52.2	lymphoblastic leukaemia, on chemotherapy	Immunocompromised		0	N	0	5	B.2.1	(76–78)
D	Male	0-1	White	1.98	ex 32 week prematurity, required inotropes and ventilation	Immunocompetent		1	Y [10]	22	7	B.1.1	(76–79)
E	Male	2-5	Other	12.4	Neuroblastoma on chemotherapy	Immunocompromised	Suspected	1	N	0	5	B.1.p16	(76–78)
F	Male	0-1	Unspecified	7.4	extreme prematurity, short gut, chronic lung disease	Likely immunocompromised	Suspected	0	N	0	2	B.2.1	(76–78)
G	Male	11-15	Unspecified	120	obese	Immunocompetent		10	Y [10]	10	10	B.1.1.7	(76–78)
H	Male	0-1	Black African	10.3	Hepatoblastoma	Immunocompromised		Unknown	N	34	13	B.1.p11	(76–78)
I	Female	0-1	Unspecified	9.85	PNET, on chemotherapy	Immunocompromised	Suspected	Asymptomatic	N	0	2	B.1.p16	(76–78)

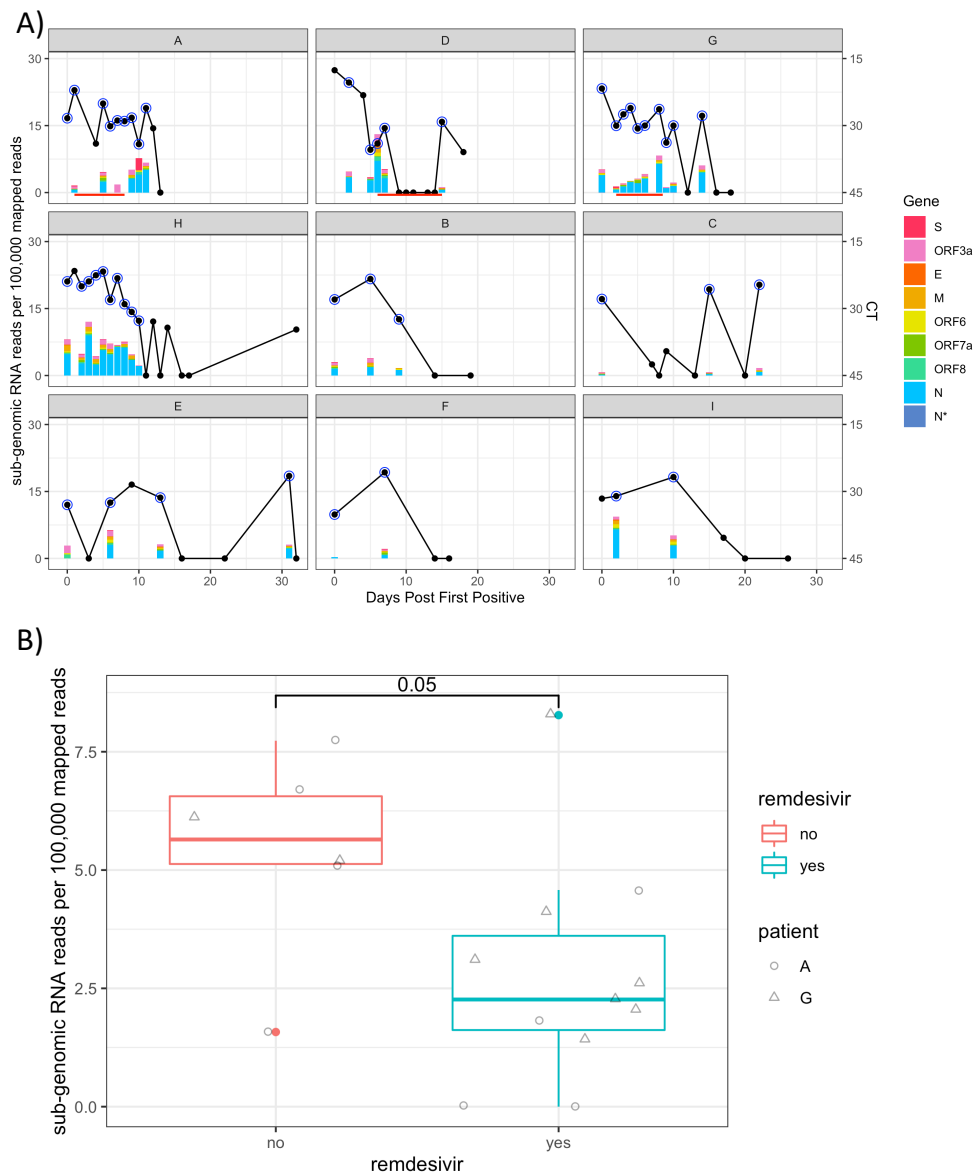
**Figure 1.** Ct trajectories of 9 patients from first day up to 40 days post first positive. One panel per patient, red line indicates remdesivir received, black dot is sample taken, blue circle indicates sample successfully sequenced. Orange dot indicates Bronchoalveolar lavage sample.



**Figure 2:** RAxML phylogenetic trees rooted at NC\_045512. **A)** Tree using consensus level sequences for Patients A-I. Boxes highlight distinct, identical sequences excluding gaps, found in patients A, H and I over time. Samples are labelled as [Patient][Time]. **B)** Tree using haplotype sequences for Patients A, B, H, and I and consensus levels sequences for Patients D-G for which no haplotypes are identified. Haplotypes defining mutations are shown along the corresponding branches. Samples are labelled as Hap [number]\_[Patient][Time].



**Figure 3:** Evaluation of effect of remdesivir on levels of sgRPHT. A) Comparison of ct values and sgRPHT over time by Patient. Stacked bars represent sgRPHT values coloured by gene. Black line represent ct values, with blue circles annotating successfully sequenced samples. Y-axis is days post first positive. B) Box-plot of sgRPHT on and off remdesivir for patients A and G. Samples from each individual are identified by their shape. Treated samples have low sgRPHT than untreated samples taken in the same time-window post first positive (Mann-Whitney-Wilcoxon test,  $p = 0.05$ ). Patient D was excluded from this comparison as no sequences were available during remdesivir treatment as viral load was below the limit of detection.



**Figure 4:** Frequency of identified haplotype over time for individual patients. **A)** Haplotype frequency over time for patients A, B, H and I. **B) right:** Phylogenetic tree of haplotypes from patients A and D, nucleotide mutation shown above each cluster *left:* Haplotype frequency over time for patients A and D. Black line is Ct value, red line indicates remdesivir received, black dot is sample taken, blue circle indicates sample successfully sequenced. Bars indicate frequency of identified haplotypes.

