

# THE LANCET Infectious Diseases

## Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Meredith L W, Hamilton W L, Warne B, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis* 2020; published online July 14. [https://doi.org/10.1016/S1473-3099\(20\)30562-4](https://doi.org/10.1016/S1473-3099(20)30562-4).

# Appendix

## Rapid implementation of SARS-CoV-2 sequencing to investigate healthcare-associated COVID-19 infections: a genomic epidemiology study

### Authors

Luke W. Meredith<sup>#\*1</sup>, William L. Hamilton<sup>#\*2,3</sup>, Ben Warne<sup>2,3</sup>, Charlotte J. Houldcroft<sup>2</sup>, Myra Hosmillo<sup>1</sup>, Aminu S. Jahun<sup>1</sup>, Martin D. Curran<sup>4</sup>, Surendra Parmar<sup>4</sup>, Laura G. Caller<sup>1,5</sup>, Sarah L. Caddy<sup>2</sup>, Fahad A. Khokhar<sup>2</sup>, Anna Yakovleva<sup>1</sup>, Grant Hall<sup>1</sup>, Theresa Feltwell<sup>2</sup>, Sally Forrest<sup>2</sup>, Sushmita Sridhar<sup>2,6</sup>, Michael P. Weekes<sup>2</sup>, Stephen Baker<sup>2</sup>, Nicholas Brown<sup>4</sup>, Elinor Moore<sup>3</sup>, Ashley Popay<sup>7</sup>, Iain Roddick<sup>7</sup>, Mark Reacher<sup>7</sup>, Theodore Gouliouris<sup>3,4</sup>, Sharon J. Peacock<sup>2,8</sup>, Gordon Dougan<sup>2</sup>, M. Estée Török<sup>#\$2,6</sup> and Ian Goodfellow<sup>#\$1</sup>

# These authors contributed equally to the work:

\* Joint first author

\$ Joint last author

### Affiliations

1. Department of Pathology, University of Cambridge, Cambridge, UK
2. Department of Medicine, University of Cambridge, Cambridge, UK
3. Cambridge University Hospitals NHS Foundation Trust, UK
4. Public Health England Clinical Microbiology and Public Health Laboratory, Cambridge, UK
5. Francis Crick Institute, London, United Kingdom
6. Wellcome Sanger Institute, Hinxton, UK
7. Field Epidemiology, Field Service, National Infection Service, Public Health England, Cambridge, United Kingdom
8. National Infection Service, Public Health England, London

### Corresponding Authors:

Dr M. Estée Török [et317@cam.ac.uk](mailto:et317@cam.ac.uk)

Professor Ian Goodfellow [ig299@cam.ac.uk](mailto:ig299@cam.ac.uk)

# Contents

Supplementary Methods .....	1
SARS-CoV-2 molecular testing.....	1
Sample and data collection .....	<b>Error! Bookmark not defined.</b>
Sequencing details .....	1
Bioinformatic analysis .....	2
Investigation of genetic clusters with zero SNP differences .....	3
Epidemiological analysis methods .....	3
Timeline plotting .....	3
Ward time and genomic cluster plots .....	3
Data and Sample Processing at CUH .....	5
Process for investigating healthcare associated COVID-19 infections .....	6
Baseline characteristics of COVID-19 patients at CUH .....	7
CUH COVID-19 infections and sequence data availability.....	9
Genome coverage plotted against Ct value .....	10
Location and frequency of SNPs across sequenced genomes.....	11
SARS-CoV-2 lineages identified over time.....	12
Phylogenetic tree and lineages of East of England genomes.....	13
Phylogenetic tree of CUH SARS-CoV-2 genomes highlighting samples taken in the Emergency Department.....	14
SNP difference matrix for CUH SARS-CoV-2 genomes .....	15
Distribution of SNP differences among CUH SARS-CoV-2 genomes .....	16
Epidemiological timelines of hospital clusters .....	21
GISAID genomes included in phylogenetic tree .....	22
Table of sequences / accession numbers.....	24

# Supplementary Methods

## SARS-CoV-2 molecular testing

Nucleic acid extraction was undertaken using the NUCLISENS easyMAG platform (Biomerieux, Marcy L-Etoile), in accordance with manufacturers' instructions. Nucleic acids were extracted from 500µL of sample, with a dilution of MS2 bacteriophage added pre-extraction to act as an internal extraction and inhibition control. The presence of SARS-CoV-2 was assessed using an in-house generated and validated one-step RT q-PCR assay that detects a 222 base-pair region of the SARS-CoV-2 RdRp genes, along with an MS2 bacteriophage internal extraction control. The RdRp gene was detected using the RdRp For primer (ATGGGTTGGGATTATCCTAAATGTGA) and the RdRp Rev primer (AGCAGTTGTGGCATCTCCTGATGAG) with a FAM labelled MGB RdRp Probe 3 (ATGCTTAGAATTATGGCCTCAC). The internal extraction control was detected using the MS2 For primer (TGCACTACCCCTCTCCGTATTCACG), the MS2 Rev primer (GTACGGGCGACCCACGATGAC) and a ROX-BHQ2 labelled MS2 probe (CACATCGATAGATCAAGGTGCCTACAAGC). Amplification reactions and detection of PCR products were performed using the Rotorgene™ PCR instrument. A typical reaction contained 400nM of For and Rev primers for the RdRp genes and 200nM of the the MS2 internal control For and Rev primer pair, along with 120nM of the RdRp and MS2 probes. TaqPath™ 1-Step RT-qPCR Mastermix (Thermo) was used. Reactions typically contained 25% extracted nucleic acid and were cycled through the following conditions: RT (25°C for 2 mins, 50°C for 15 mins, 95°C for 2 mins) followed by 45 cycles of (95°C for 3 secs and 60°C for 30 secs) acquiring on FAM and ROX on the Rotor-Gene Q real-time PCR instrument. Samples that generated a Ct value ≤36 were considered positive. Samples and negative control (molecular grade water) were individually spiked with MS2 bacteriophage internal control (4600 pfu per extraction) prior to nucleic acid extraction to identify any inhibitors or extraction issues. Positive control material, BetaCoV/England/02/2020, was obtained from PHE Colindale and was essentially purified virus RNA diluted down to give a cycle threshold value of 26-28. Negative controls included extracted molecular grade water.

## Sequencing details

Samples were sequenced using Nanopore technology following the ARTICnetwork V3 protocol (<https://dx.doi.org/10.17504/protocols.io.bbmuik6w>) and assembled using the ARTICnetwork assembly pipeline (<https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>). Median genome depth of coverage was 6,612x across all 747 genomes. 14 samples in our dataset were also sequenced with Illumina technology at the Wellcome Sanger Institute as part of COG-UK. There was 100% concordance in called nucleotides between sample pairs. Four genomes differed because of base pairs called in the Illumina data that were missing in the Nanopore sequences. The accession numbers of the samples included in this study are available in Appendix pp 24-28.

## Bioinformatic analysis

Consensus fasta sequence quality control cutoffs were: size >29Kb, N count <2990 (~10%). After QC filtering, de-duplication and matching with metadata, the first sample set analysed comprised 197 genomes collected up to 10th April 2020; set 2 had 444 genomes up to 15th April, and set 3 (presented here) had 747 genomes up to 24th April. 30 reference genomes were added to the sample sets downloaded from GISAID (<https://www.gisaid.org/>; Appendix pp 22-23). The reference genomes were chosen to represent the major branches of the global phylogenetic tree as visualised in Nextstrain (<https://nextstrain.org/>) to provide broader context, including a sample from December 2019 collected in Wuhan, China, used to root the tree. Multiple sequence alignment was performed using MAFFT (v 7.458) with default settings, command:

```
/PATH/mafft" --retree 2 --inputorder "multi_fasta.fasta" > "aligned_multi_fasta
```

The alignment was manually inspected using AliView. Maximum likelihood trees were produced using IQ-TREE software<sup>15</sup> for all samples passing QC filters and the subset of samples from CUH (n=299 for this dataset). Initial tests with the ModelFinder Plus option<sup>32</sup>, which selects the optimal nucleotide substitution model out of over 200 options (<http://www.iqtree.org/doc/Substitution-Models>), consistently identified GTR+F+I as the best model. Therefore from 24th April (including analysis presented here) we specified GTR+F+I.

Command using ModelFinder Plus:

```
/PATH/iqtree -s aligned_filtered_multi_fasta -m MFP
```

Command with GTR+F+I model specified:

```
/PATH/iqtree -s aligned_filtered_multi_fasta -m GTR+F+I
```

Trees were manually inspected in FigTree, rooted on the 2019 Wuhan sample (EPI\_ISL\_402123), ordered by descending node and exported as Newick files. Trees were visualised in online software Microreact<sup>16</sup> in a private account to explore relationships between wards and clinico-epidemiological questions for our weekly reports. Further visualisations were produced in R using the packages *Ape*<sup>33</sup> (v 5.3) and *ggtree*<sup>17</sup> (v 2.0.4).

A SNP difference matrix was produced from the multiple sequence alignment using the *snp-dists* package (v 0.7.0; <https://github.com/tseemann/snp-dists>; installed into a conda environment), command:

```
snp-dists -c aligned_filtered_multi_fasta.aln > snp_dist_matrix.csv
```

The matrix was exported as .csv and manipulated in R using the *Matrix* and *tidyverse* packages for ward and pairwise SNP comparisons and plotted using the *ggplot2* (v 3.3.0) package. A heatmap was produced in python using the *seaborn* (v 0.10.0) *clustermap* function. To identify clusters with zero SNP differences we initially used the *scipy.cluster.hierarchy* functions *linkage* and *fcluster* (*scipy* v 1.4.1), with additional samples in complete linkage (zero SNP differences between all members of the cluster) identified using a custom R script that searched for zero SNP differences between pairwise sample comparisons and kept the largest groups containing each sample. Clusters were named in descending size order and linked with sample metadata and lineage data.

## Investigation of genetic clusters with zero SNP differences

Patient records from each case within a putative genomic cluster were manually reviewed in detail by authors BW, WLH and MET and assigned a score of 1 (strong evidence supporting a recent linked transmission chain, e.g. patients co-located on the same ward becoming positive within the incubation period of the virus), 2 (a plausible transmission chain is present e.g. patients becoming positive while located in nearby wards within the hospital but who did not appear to be in direct contact), and 3 (no evidence of any connections between cases) – see Appendix pp 17 – 20 for further details.

## Epidemiological analysis methods

### Timeline plotting

Space time relationships between patients were plotted using patient specific time lines by exporting the bed and ward admission dates, dates of transfer, dates of discharge or date of death obtained from hospital information system (EPIC Systems Corporation, Verona, USA) and importing them into a cloud-based timeline plotter application (Cluster Track, Camart Ltd, Cambridge available at Clusterack.com). Earliest positive specimen date for COVID19 was obtained from the laboratory records and date of onset of symptoms from the clinical records and uploaded.

The application aided visualisation of ward and time relations by assigning unique colours to wards and then ward presence by date along the x axis shown in days, such that a solid timeline bar by colour and by date permitted the visualisation of the location of each patient over time. The positive specimen date for COVID 19, genomic cluster, death and discharge were each overlaid on the patient timeline using standard visual representation built into the application. Visualisation was aided by using the sort command within the application on admission date, earliest positive specimen date, or first ward to which admitted. Separate plots of subset of the total cases were created to provide clearer visualisation when needed

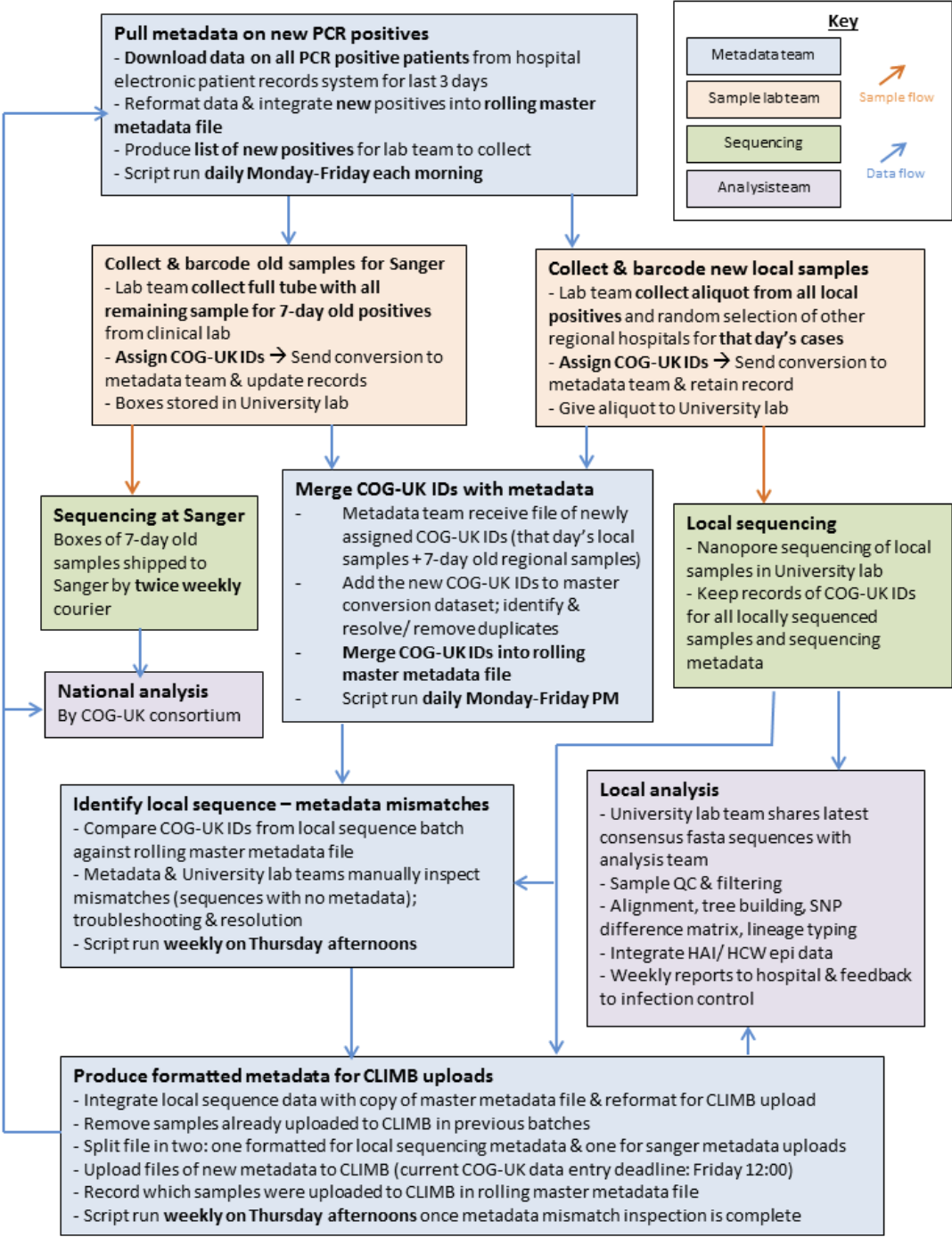
### Ward time and genomic cluster plots

A clustering and network analysis function was used in the Cluster track application in which an algorithm links patients with admission days to the same ward on the same date and displays a network diagram to indicate these overlapping cases.

More advanced space time clustering was undertaken by exporting these timeline data sets into an SQL database running a more advanced clustering algorithm in which time parameters were set for the presumed susceptible, infectious and non-infectious/ recovered intervals counting from the earliest positive specimen date. The algorithm identified and linked cases in which two or more patients had an overlap on the same ward of the time interval of infectiousness of an earlier case with the interval of susceptibility in a later case or cases. Links continue to be made until no further overlaps of the infectious interval in an earlier case occurred with the interval of susceptibility in a later case on the same ward: this ended the space time cluster.

The cluster diagrams of the space time clustered cases were reviewed. Cases within the same space time cluster that belonged to the same genomic cluster were deemed to be supportive of a recent transmission event.

# Data and Sample Processing at CUH

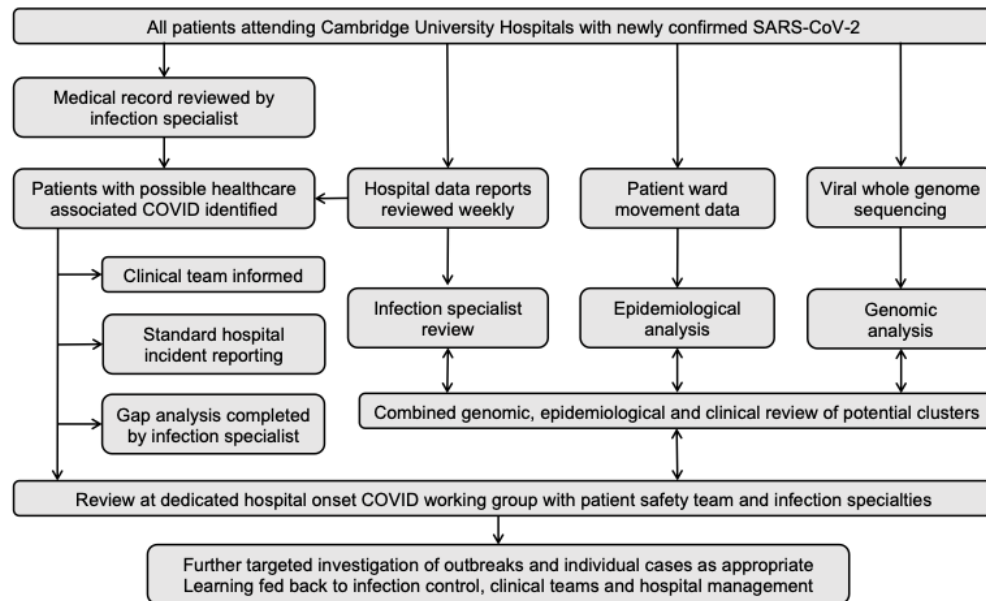


HAI = Hospital Acquired Infection  
 HCW = Healthcare Worker

Flow diagram representing sample and metadata flow between clinical diagnostics and sequencing centres.



## Process for investigating healthcare associated COVID-19 infections



Conceptual flow diagram shows investigation process for healthcare associated COVID-19 infections at Cambridge University Hospitals NHS Foundation Trust. Review meetings took place weekly.

## Baseline characteristics of COVID-19 patients at CUH

<b>Baseline characteristics</b>	<b>Singletons</b>	<b>Clusters</b>	<b>No sequence</b>	<b>Total</b>
Number of patients	126	136	112	374
Age in years, mean (range) Age in years, median (IQR)	62 (0-97) 64 (48-77)	68 (10-98) 71 (57-83)	62 (0-93) 64 (51-77)	64 (0-98) 67 (51-79)
Male sex	78 (61.9%)	82 (60.3%)	73 (65.2%)	233 (62.3%)
Female sex	48 (38.1%)	54 (39.7%)	39 (34.8%)	141 (37.7%)
Ethnicity – White	86 (68.3%)	104 (76.5%)	79 (70.5%)	269 (71.9%)
Ethnicity – Black, Asian and minority ethnic	7 (5.6%)	8 (5.9%)	12 (10.7%)	27 (7.2%)
Ethnicity – not stated/missing	33 (26.2%)	24 (17.7%)	21 (18.8%)	78 (20.9%)
<b>Co-morbidities</b>				
Hypertension	35 (27.8%)	45 (33.1%)	37 (33.0%)	117 (31.3%)
Ischaemic heart disease	13 (10.3%)	26 (19.1%)	15 (13.4%)	54 (14.4%)
Cardiac failure	7 (5.6%)	10 (7.4%)	7 (6.3%)	24 (6.4%)
Asthma	12 (9.5%)	15 (11.0%)	20 (17.9%)	47 (12.6%)
Chronic obstructive pulmonary disease	11 (8.7%)	13 (9.6%)	7 (6.3%)	31 (8.3%)
Diabetes mellitus	18 (14.3%)	38 (27.9%)	22 (19.6%)	78 (20.9%)
Chronic kidney disease	14 (11.1%)	19 (14.0%)	4 (3.6%)	37 (9.9%)
Hepatic cirrhosis	7 (5.6%)	7 (5.2%)	2 (1.8%)	16 (4.3%)

Dementia	7 (5.6%)	20 (14.7%)	5 (4.5%)	32 (8.6%)
Obesity	20 (15.9%)	20 (14.7%)	14 (12.5%)	54 (14.4%)
<b>Classification of infection</b>				
Community onset, community associated	100 (79.4%)	76 (55.9%)	87 (77.7%)	263 (70.3%)
Community onset, suspected healthcare-associated	8 (6.4%)	12 (8.8%)	12 (10.7%)	32 (8.6%)
Hospital onset, indeterminate healthcare-associated	5 (4.0%)	4 (2.9%)	4 (3.6%)	13 (3.5%)
Hospital onset, suspected healthcare-associated	1 (0.8%)	10 (7.4%)	3 (2.7%)	14 (3.7%)
Hospital onset, healthcare-associated	11 (8.7%)	27 (19.9%)	5 (4.5%)	43 (11.5%)
Healthcare worker	1 (0.8%)	7 (5.2%)	1 (0.9%)	9 (2.4%)
<b>Outcome</b>				
Admission to hospital	115 (91.3%)	129 (94.9%)	102 (91.1%)	346 (92.5%)
Admission to critical care	37 (29.4%)	22 (16.2%)	15 (13.4%)	74 (19.8%)
Died as an inpatient	33 (26.2%)	32 (23.5%)	13 (11.6%)	75 (20.1%)

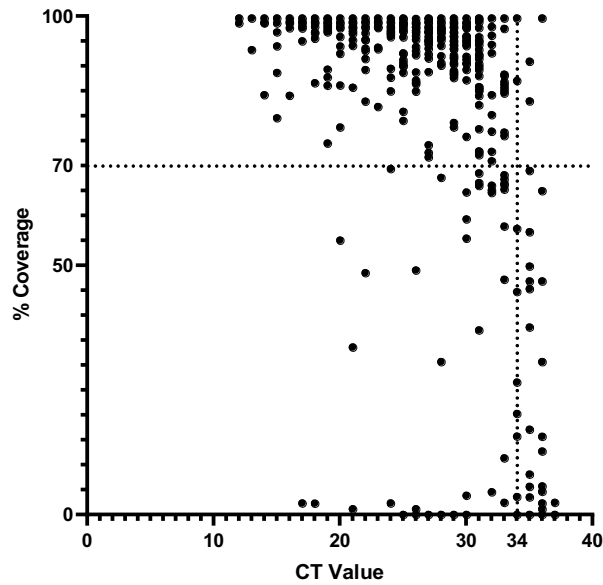
Baseline characteristics of confirmed COVID-19 patients at CUH (with confirmed results between 10<sup>th</sup> March to 24<sup>th</sup> April, excluding 37 healthcare workers diagnosed as part of staff screening. Genomic clusters were defined as 2 or more identical virus. Genomic singletons had unique genomes in the dataset.

## CUH COVID-19 infections and sequence data availability

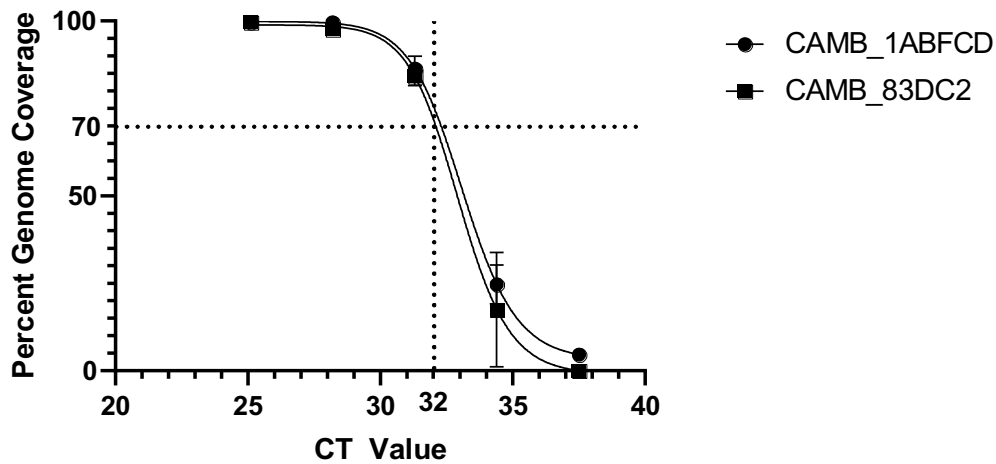
<b>Classification of infection</b>	<b>No.</b>	<b>No. with available sequence (%)</b>
Community onset, community associated	263	176 (66.9%)
Community onset, suspected healthcare-associated	32	20 (62.5%)
Hospital onset, indeterminate healthcare-associated	13	9 (69.2%)
Hospital onset, suspected healthcare-associated	14	11 (78.6%)
Hospital onset, healthcare-associated	43	38 (88.4%)
Healthcare worker	9	8 (88.9%)
Total	374	262 (70.1%)

Table shows breakdown of COVID-19 infection classification at CUH and availability of SARS-CoV-2 sequences for analysis.

## Genome coverage plotted against Ct value

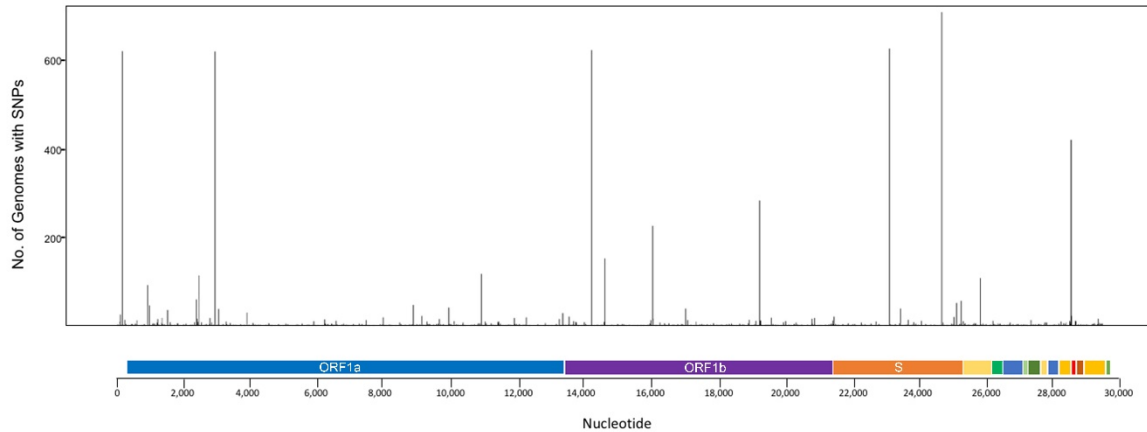


Ct value plotted against the percent of SARS-CoV-2 genome sequenced prior to internal screening and for which a Ct value is available (N=947). Median Ct value of samples failing 70% coverage threshold is 34 (N=85).



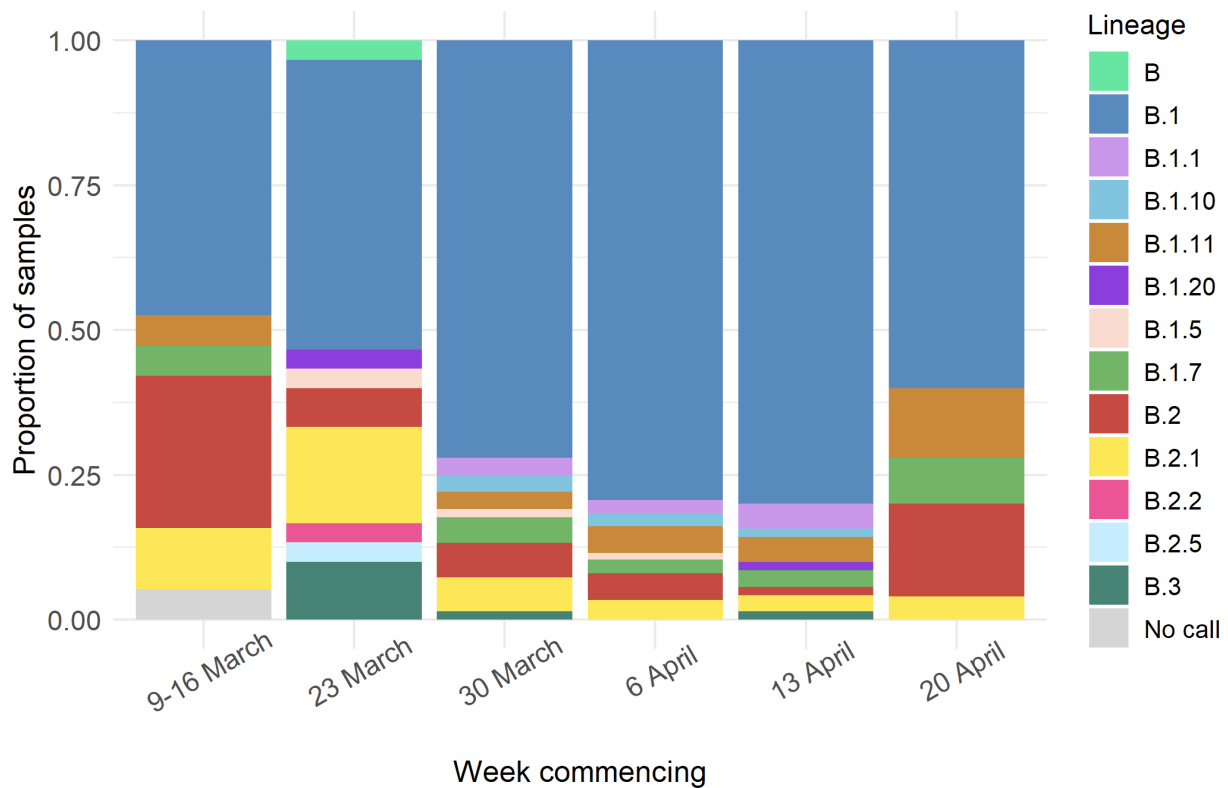
Correlation between Ct and Percent Genome Coverage for two biological samples diluted 1:9 from a Ct of 25 to 37. Samples were sequenced in duplicate. The effective Ct value yielding 70% genome coverage averaged  $32.19 \pm 0.14$  (n=2, SD).

## Location and frequency of SNPs across sequenced genomes



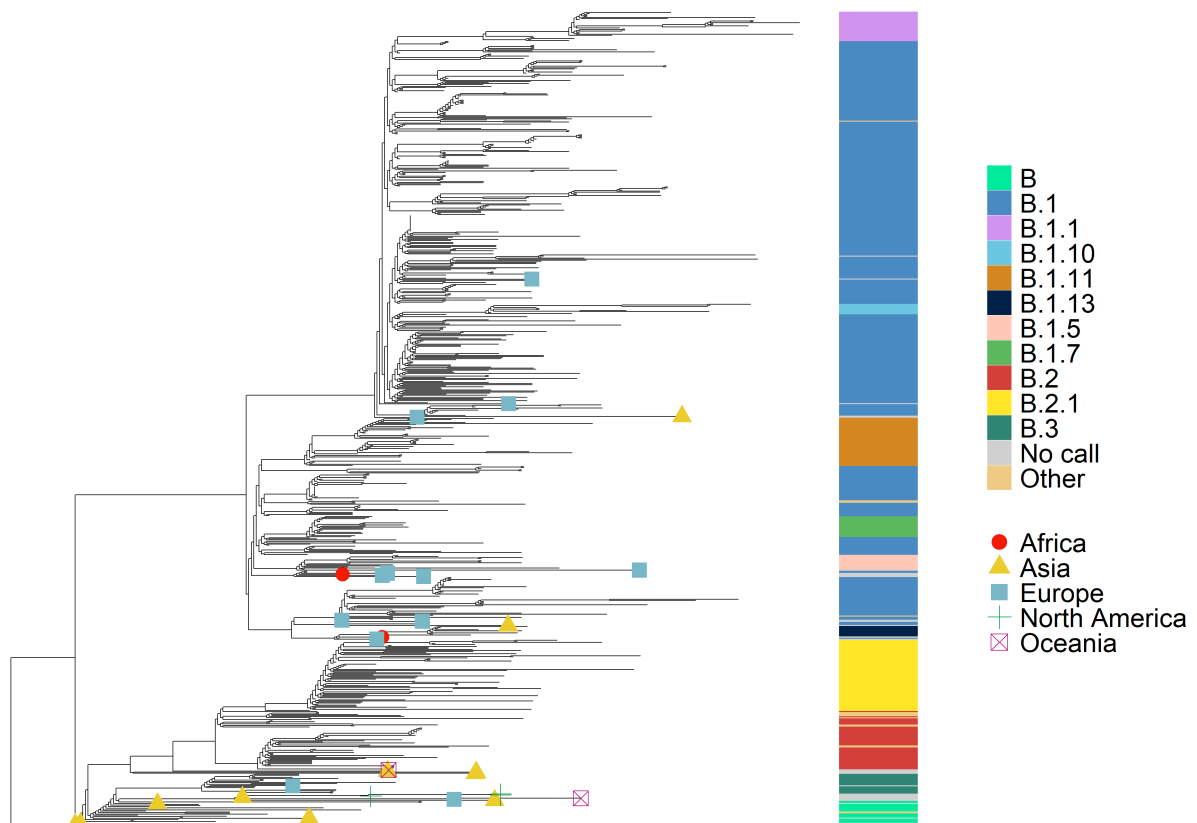
Cumulative number and location of SNPs compared to the original Wuhan strain (Accession No. MN908947) observed across 747 genomes sequenced in East of England. This shows the total occurrence of SNPs, 10,536nt across 22,337,541nt sequenced (0.005%) occurring at 1,196 positions. Of the 1,196 positions, 1,192 SNPs were found to be single SNPs, while 4 sites had 2 SNPs. 5 common SNPs were found in the majority of sequenced genomes (A23403G, C14408T, C241T, C3037T, T deletion at 24981) while G28881A, G28882A, G28883 were also found in ~50% of samples. These are not unique mutations and have been observed in other cases in the global NextStrain analysis.

## SARS-CoV-2 lineages identified over time



In the weeks commencing 9 and 13 March 2020 lineages (and descendent lineages) of B, B.1, B.2, B.3 and B.8 were present in the EoE (amalgamated here as there were only 2 samples for week commencing 9 March). Diversification of lineages already present from earlier weeks was seen over time, with the detection of descendants of lineages B.1 and B.2, but no new lineages emerged during this period, likely an impact of lockdown measures preventing new viruses being introduced from other regions. Changes in lineage frequency may be stochastic due to changes in the available sample size during each week of the sampling period. Lineage B.8 was only detected in the week commencing 16 March 2020. Lineage B.4 viruses (associated with export from Iran) were not seen in our sample set. Lineage A viruses (or A descendants), most commonly reported in China, USA, South Korea and Australia, were not detected in our EoE samples.

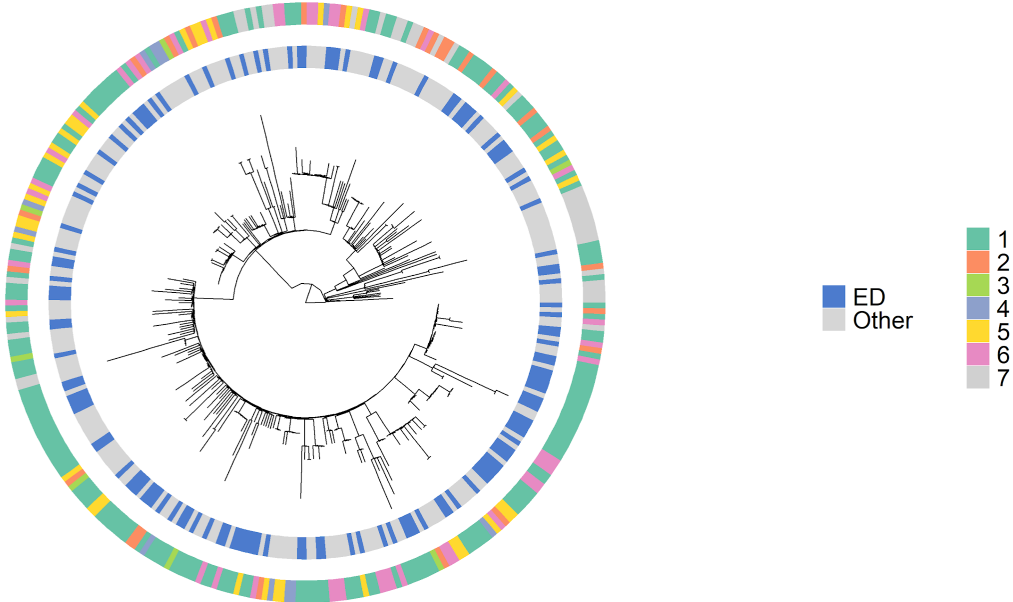
## Phylogenetic tree and lineages of East of England genomes



Phylogenetic tree of 747 East of England SARS-CoV-2 genomes and 30 reference genomes used to provide further genetic and geographic context. The reference genomes are highlighted with coloured tips and are the same used in Figure 4. As with Figure 4, the tree is rooted on a December 2019 sample from Wuhan, China, with older samples from Asia represented at the base of the tree as expected. The lineages are indicated by the colour bar.

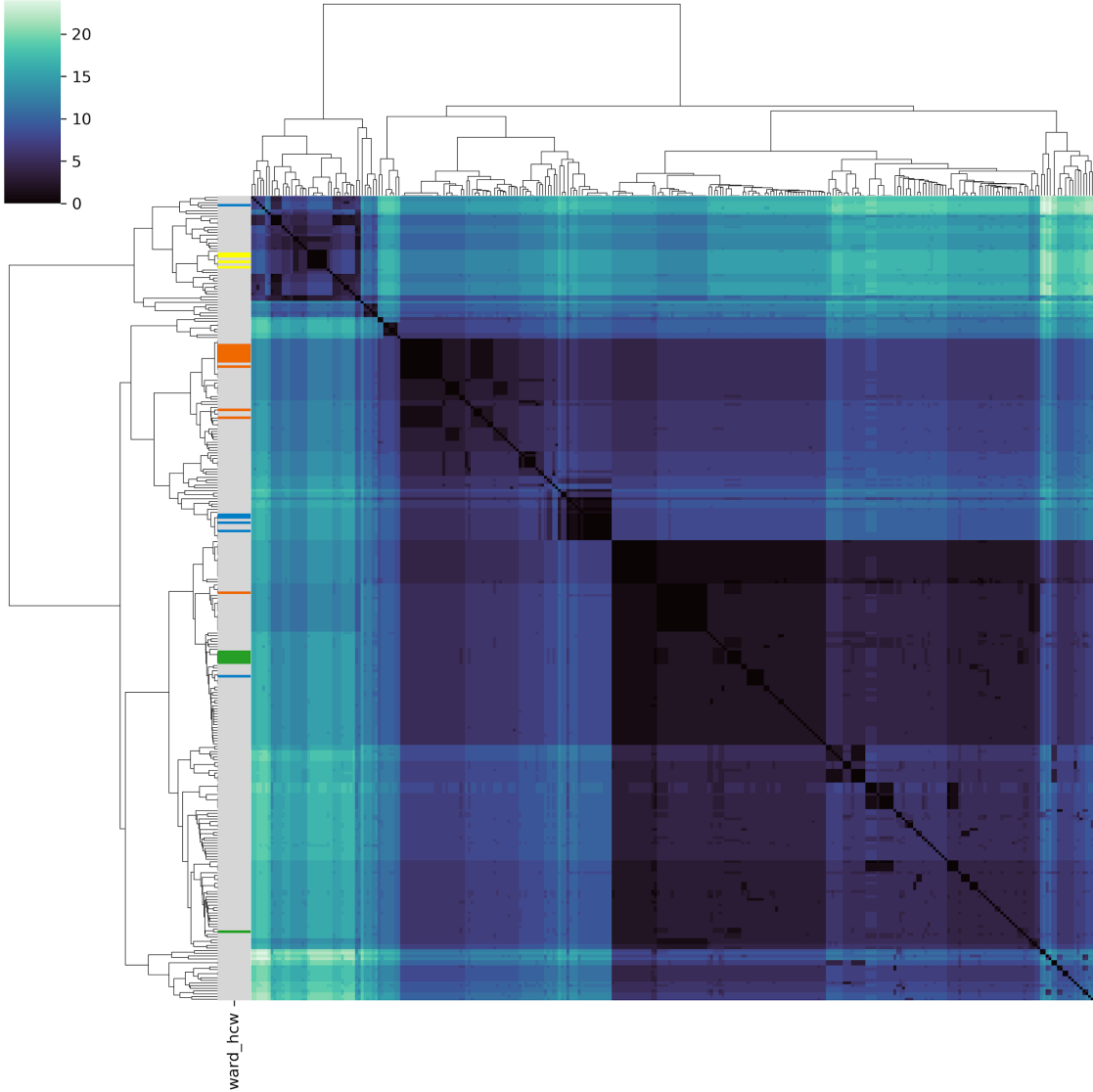


# Phylogenetic tree of CUH SARS-CoV-2 genomes highlighting samples taken in the Emergency Department



Phylogenetic tree of 299 CUH SARS-CoV-2 genomes and 30 reference genomes. The inner ring shows emergency department (ED) samples in blue and samples collected from all other sites in grey. The outer ring shows the different classifications of infection: 1. Community onset, community associated; 2. Community onset, suspected healthcare-associated; 3. Hospital onset, indeterminate healthcare-associated; 4. Hospital onset, suspected healthcare-associated; 5. Hospital onset, healthcare-associated; 6. Healthcare worker; 7. Unable to determine/ data missing.

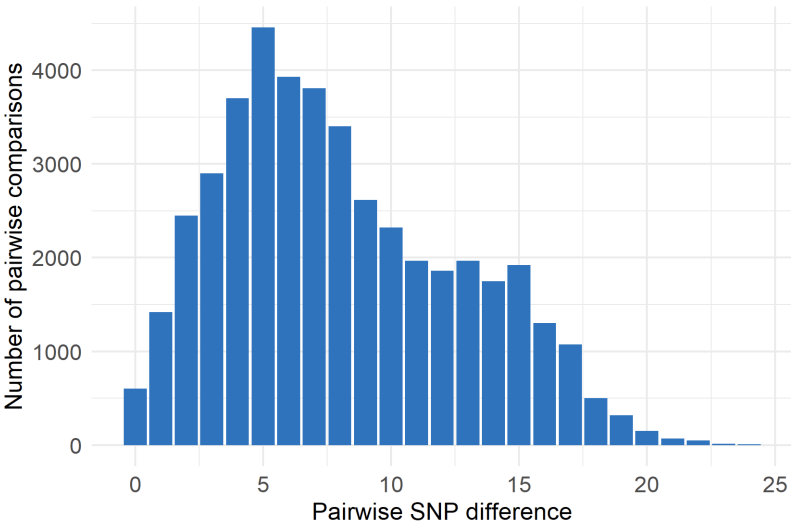
# SNP difference matrix for CUH SARS-CoV-2 genomes



SNP difference matrix for 299 SARS-CoV-2 genomes from CUH. Darker colouring indicates more similar genomes. Several clusters of identical viruses are present in the dataset, as discussed in main text. The left-hand bar shows wards A, B, C and the dialysis unit highlighted in colour and other wards in grey

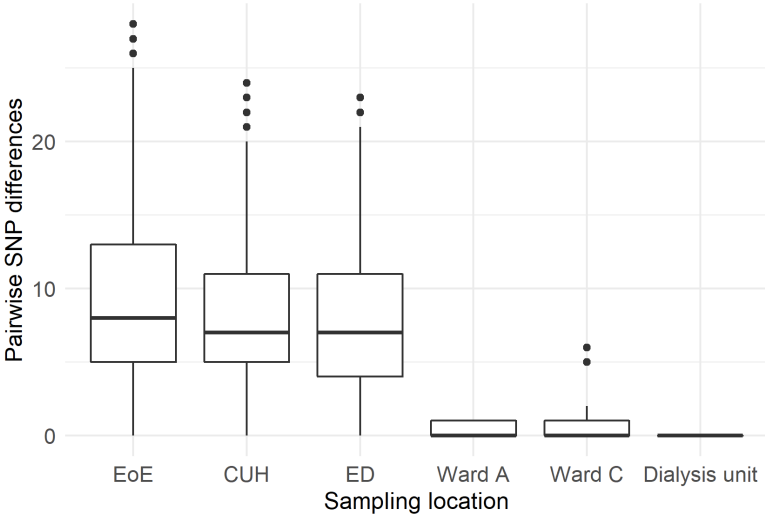
# Distribution of SNP differences among CUH SARS-CoV-2 genomes

## Frequency distribution of pairwise SNP differences between CUH SARS-CoV-2 genomes



Distribution of pairwise SNP differences for 299 CUH samples. The total number of pairwise comparisons is 44,551. The median difference was 8 SNPs (range 0 to 24 SNPs). 4.5% of genomes were identical or 1-SNP different.

## Box plot of SNP differences between SARS-CoV-2 genomes from selected sampling locations



Box plot of SNP differences between SARS-CoV-2 genomes at CUH within different sampling locations. The number of SNP differences was very low on certain wards (0 to 1 SNPs) compared with the emergency department (ED), the Trust (CUH) and the East of England (EoE) as a whole, consistent with shared recent transmission events on these wards (discussed in main text).

## Epidemiological analysis of clusters of identical SARS-CoV-2 genomes

No.	Size	CAIs	Possible HAIs*	HAIs	HCWs	First Case	Last Case	Cases with strong epi links	Cases with possible epi links	Cases with no epi links	Notes
1	18	12	1	0	5	5/4/20	21/4/20	10	6	2	Nine patients were resident in care home A. One patient works as a carer in the same home. Another patient also works as a carer in an unspecified home. Three of the HCWs were paramedics. The other two HCWs work in different clinical areas, but both live with paramedics. Two patients had no identifiable contact with the other cases in this cluster.
2	16	13	0	1	2	20/3/20	20/4/20	2	0	14	No identified epidemiological link between the 14 patients. One HCW had direct contact with two of these patients close to their admission. The second worked on the same ward as one patient at the time of their admission (ward E).
3	15	4	4	5	2	20/3/20	16/4/20	11	1	3	Eight patients swabbed during a suspected outbreak on ward C. The first patient to be swabbed was recently discharged from ward D and then re-admitted to ward E. The two HCWs worked on ward D; one had direct contact with the first patient to be swabbed in this cluster. Two patients, distinct from the larger ward outbreak, co-habit (spouses). The final patient has no identifiable association with any other case in this cluster.
4	12	1	3	3	5	1/4/20	20/4/20	9	2	1	Four patients swabbed during a suspected outbreak on ward B; a fifth patient was discharged from ward B and re-admitted with COVID within 2 weeks of discharge. Two HCWs work on ward F; one of these HCWs also work on ward B, alongside two of the other HCWs. One patient lives with a HCW (unknown clinical area). There are no

											identified associations with the remaining patient or the fifth HCW.
<b>5</b>	7	0	6	0	1	1/4/20	22/4/20	6	1	0	All six patients receive dialysis at the same unit. The HCW works on two wards on which three of these patients were admitted.
<b>6</b>	6	3	0	0	3	7/4/20	17/4/20	5	1	0	Three patients resident in care home B; one of these patients admitted to ward F. Two HCWs work on ward F, including one with direct contact with this patient. The third HCW works on ward E. The isolates from clusters 6 and 12 diverge by 1 SNP, but share epidemiological links of both patients and HCWs.
<b>7</b>	6	4	2	0	0	30/3/20	19/4/20	0	6	0	Two patients live in separate care homes. The third patient works in an unspecified care home. The fourth is a community carer. The fifth lives with a carer working in an unspecified care home. No patients have other identified associations. The final patient has no association with any of the patients, but was cared for on neighbouring wards with some shared staff.
<b>8</b>	5	0	3	2	0	27/3/20	11/4/20	5	0	0	Five patients all swabbed on ward A.
<b>9</b>	5	0	4	0	1	20/3/20	16/4/20	4	1	0	Three patients swabbed during a suspected outbreak on ward D. HCW works on ward D. The first patient in this cluster was recently discharged from ward E and readmitted to ward E. Wards D and E are in the same department and share some staff
<b>10</b>	5	1	0	2	2	31/3/20	17/4/20	4	1	0	Two patients were on ward I at the time of swabbing, with overlapping admissions. One patient is the husband of one of the HCWs. The HCWs have no known association with each other or the patients.
<b>11</b>	5	1	0	3	1	12/4/20	20/4/20	5	0	0	The two patients were co-located on ward J. One of these patients was co-located with the third patient on ward K. The two HCWs work on ward J, which has shared staff with ward K.
<b>12</b>	4	2	0	0	2	4/4/20	17/4/20	4	0	0	One patient resident in care home B, admitted to ward G. Second patient is a carer in care home B. One HCW works

											on wards G and E. The second HCW works on ward E. The isolates from clusters 6 and 9 diverge by 1 SNP, but share epidemiological links of both patients and HCWs.
<b>13</b>	4	2	1	1	0	15/3/20	15/4/20	0	0	4	No identified associations between the four patients
<b>14</b>	3	3	0	0	0	11/4/20	20/4/20	0	2	1	One patient is a resident in a specialist dementia care home. A second patient works in a specialist dementia care home, but it is unclear whether this is the same home as the first patient. The third patient has no other identified associations with the other patients.
<b>15</b>	3	0	1	0	2	2/4/20	14/4/20	3	0	0	The patient was recently discharged from ward I and readmitted to ward F with documented direct contact with both HCWs.
<b>16</b>	3	1	0	0	2	16/4/20	21/4/20	0	3	0	Two HCWs, working in separate wards in the same department. The partner of one of these HCWs works in an unspecified care home. The patient in this cluster was admitted from a care home.
<b>17</b>	3	1	0	0	2	15/4/20	18/4/20	2	1	0	The two HCWs live in the same home, with a care assistant in an unspecified care home. The patient was admitted from a care home.
<b>18</b>	3	0	0	2	1	27/3/20	12/4/20	2	0	1	Two patients were co-located on ward C. No direct contact documented between the HCW and the two patients.
<b>19</b>	3	2	0	1	0	8/4/20	8/4/20	2	1	0	Two patients live together (spouses) in their own home and have carers four times a day. The third patient lives in the same village and has carers twice a day.
<b>20</b>	3	3	0	0	0	13/3/20	21/4/20	2	0	1	Two patients live in the same hostel. No known association with the third patient.
<b>21</b>	2	1	1	0	0	4/4/20	5/4/20	2	0	0	Two patients co-located on ward H.
<b>22</b>	2	1	0	0	1	19/4/20	20/4/20	0	0	2	The patient works in a local community hospital. The HCW works in a rehabilitation unit. There are no other known epidemiological associations.
<b>23</b>	2	1	0	1	0	3/4/20	3/4/20	0	2	0	Two patients briefly co-located within 24 hrs of testing, likely insufficient duration for transmission. One patient was resident in a care home; the second patient was on a rehabilitation ward prior to swabbing

<b>24</b>	2	2	0	0	0	22/3/20	28/3/20	0	0	2	No identified associations between the two patients
<b>25</b>	2	0	0	2	0	29/3/20	7/4/20	2	0	0	Two patients co-located on ward J
<b>26</b>	2	2	0	0	0	1/4/20	2/4/20	2	0	0	Two patients co-habiting (siblings)
<b>27</b>	2	2	0	0	0	19/3/20	19/3/20	0	0	2	No identified associations between the two patients
<b>28</b>	2	2	0	0	0	1/4/20	3/4/20	2	0	0	Two patients are mother and son (not co-habiting)
<b>29</b>	2	2	0	0	0	2/4/20	5/4/20	0	2	0	Both patients live in the same village.
<b>30</b>	2	2	0	0	0	20/3/20	27/3/20	0	2	0	Both patients live in the same village.
<b>31</b>	2	0	0	2	0	30/3/20	8/4/20	2	0	0	Two patients co-located on ward L.
<b>32</b>	2	0	0	2	0	7/4/20	8/4/20	2	0	0	Two patients co-located on ward I.
<b>33</b>	2	2	0	0	0	7/4/20	9/4/20	2	0	0	Two patients co-habiting.
<b>34</b>	2	2	0	0	0	17/3/20	11/4/20	0	0	2	No identified associations between the two patients
<b>35</b>	2	2	0	0	0	4/4/20	16/4/20	2	0	0	Two patients resident in a care home.
<b>Tot</b>	<b>159</b>	<b>75</b>	<b>26</b>	<b>27</b>	<b>30</b>	<b>13/3/20</b>	<b>22/4/20</b>	<b>92</b>	<b>32</b>	<b>35</b>	

Descriptions of the 35 clusters of genomically identical viruses (zero SNP differences) in this study.

\*Possible HAIs = patients swabbed 2-14 days from admission, or patients swabbed <2 days from admission who have had healthcare contact in the 2 weeks prior to admission (categories 2-4 in table 1, main paper).

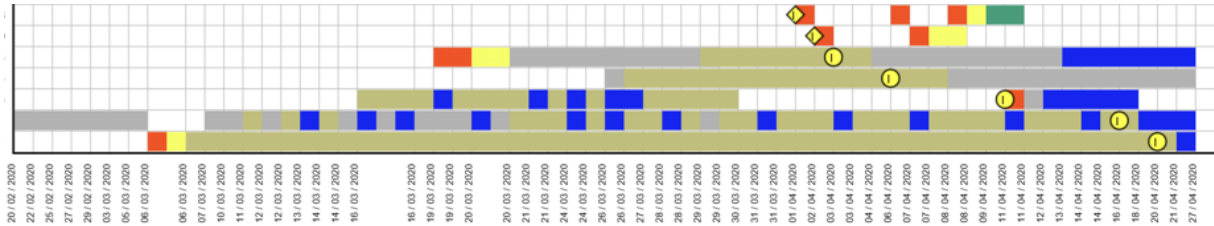
Strong epidemiological link defined as either: patient co-location in the same clinical area within the incubation period of the virus (for hospital-acquired cases); cases with the same residential address (community acquired cases); patients working in social care in the same named care home as a patient resident in this home; HCWs working in the same clinical area as other HCWs or patients.

Plausible epidemiological links defined as: patients working in social care in an unnamed care home in the same genomic cluster as a patient resident in a care home; HCWs working on different clinical areas within the same hospital department as other HCWs or patients; patients temporally co-located on neighbouring wards or clinical areas within the same department.

# Epidemiological timelines of hospital clusters

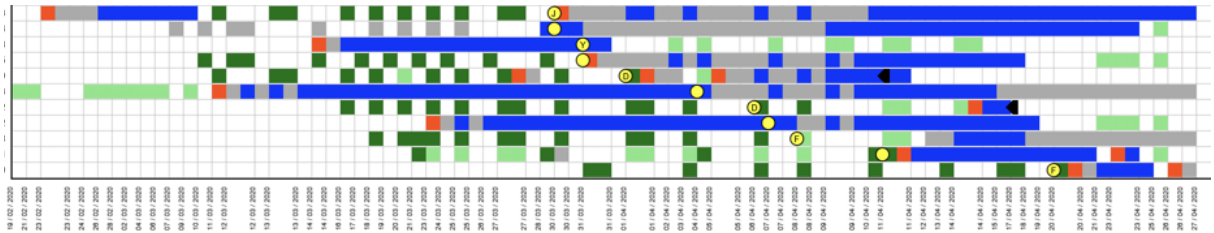
Timeline plots generated using Cluster Track (detailed in Appendix p 3).

## Hospital cluster 2 (Ward B)



Four transplant patients on ward B (shown in khaki) were diagnosed with COVID-19 infection between 3 and 20 April 2020. A fifth patient, who had been recently discharged from the ward, presented to the ED with COVID-19 infection. The sample dates are shown in yellow circles (patients) and diamonds (HCW). Genomic analysis revealed that all 5 patients had identical genomes. Two HCW were found to have identical genomes in the same cluster as the ward B cases; one had worked on ward B and had professional contact with the other HCW. The renal ward is shown in blue, the emergency department in red and the admissions unit in yellow. Other wards are shown in grey.

## Dialysis unit cluster



Six patients with end-stage renal failure were diagnosed with COVID-19 between 1 and 20 April 2020, testing positive in several locations including ED and an acute admissions ward. The sample dates are shown in yellow circles. Their viral genomes were identical, and epidemiological investigation revealed they dialysed at the same outpatient dialysis unit. This suggests linked recent transmission of community-onset healthcare-associated infections. Black triangles indicate patient deaths. The darker green blocks represent the dialysis unit with suspected transmission; the light green and grey blocks represent different dialysis units. The renal ward is shown in blue and the emergency department in red. Other wards are shown in grey.



## GISAID genomes included in phylogenetic tree

<b>GISAID ID</b>	<b>Country of origin</b>
EPI_ISL_402123*	China
EPI_ISL_406716	China
EPI_ISL_406801	China
EPI_ISL_414598	Spain
EPI_ISL_416396	China
EPI_ISL_416757	France
EPI_ISL_417969	Spain
EPI_ISL_418034	USA
EPI_ISL_419228	China
EPI_ISL_419232	China
EPI_ISL_420064	France
EPI_ISL_421905	UK
EPI_ISL_422024	UK
EPI_ISL_423034	Netherlands
EPI_ISL_424657	Belgium
EPI_ISL_426019	UK
EPI_ISL_427119	Australia
EPI_ISL_427144	Australia
EPI_ISL_427322	Russia
EPI_ISL_427391	Turkey
EPI_ISL_427441	USA
EPI_ISL_428482	India
EPI_ISL_428848	Singapore
EPI_ISL_428857	The Gambia

EPI_ISL_429061	USA
EPI_ISL_429175	Thailand
EPI_ISL_429177	Thailand
EPI_ISL_429206	Switzerland
EPI_ISL_429259	DRC
EPI_ISL_429773	Luxembourg

\*Sample from China used to root the tree









GISAID Accession ID	Virus name	COG-UK	GISAID Accession ID	Virus name	COG-UK
EPI_ISL_433772	hCoV-19/England/CAMB-787B3/2020	CAMB-787B3	Pending	hCoV-19/England/CAMB-81FDD/2020	CAMB-81FDD
EPI_ISL_433773	hCoV-19/England/CAMB-787C2/2020	CAMB-787C2	Pending	hCoV-19/England/CAMB-81FEC/2020	CAMB-81FEC
EPI_ISL_433774	hCoV-19/England/CAMB-787D1/2020	CAMB-787D1	Pending	hCoV-19/England/CAMB-81FFB/2020	CAMB-81FFB
EPI_ISL_433775	hCoV-19/England/CAMB-787E0/2020	CAMB-787E0	Pending	hCoV-19/England/CAMB-82006/2020	CAMB-82006
EPI_ISL_433776	hCoV-19/England/CAMB-7880B/2020	CAMB-7880B	Pending	hCoV-19/England/CAMB-82015/2020	CAMB-82015
EPI_ISL_433777	hCoV-19/England/CAMB-7881A/2020	CAMB-7881A	Pending	hCoV-19/England/CAMB-82024/2020	CAMB-82024
EPI_ISL_433778	hCoV-19/England/CAMB-78829/2020	CAMB-78829	Pending	hCoV-19/England/CAMB-82033/2020	CAMB-82033
EPI_ISL_433779	hCoV-19/England/CAMB-78838/2020	CAMB-78838	Pending	hCoV-19/England/CAMB-82042/2020	CAMB-82042
EPI_ISL_433780	hCoV-19/England/CAMB-78847/2020	CAMB-78847	Pending	hCoV-19/England/CAMB-82051/2020	CAMB-82051
EPI_ISL_433781	hCoV-19/England/CAMB-78865/2020	CAMB-78865	Pending	hCoV-19/England/CAMB-82060/2020	CAMB-82060
EPI_ISL_433782	hCoV-19/England/CAMB-78874/2020	CAMB-78874	Pending	hCoV-19/England/CAMB-8207F/2020	CAMB-8207F
EPI_ISL_433783	hCoV-19/England/CAMB-78883/2020	CAMB-78883	Pending	hCoV-19/England/CAMB-8208E/2020	CAMB-8208E
EPI_ISL_433784	hCoV-19/England/CAMB-78892/2020	CAMB-78892	Pending	hCoV-19/England/CAMB-8209D/2020	CAMB-8209D
EPI_ISL_433785	hCoV-19/England/CAMB-788A1/2020	CAMB-788A1	Pending	hCoV-19/England/CAMB-820AC/2020	CAMB-820AC
EPI_ISL_433786	hCoV-19/England/CAMB-788B0/2020	CAMB-788B0	Pending	hCoV-19/England/CAMB-820CA/2020	CAMB-820CA
EPI_ISL_433787	hCoV-19/England/CAMB-788CF/2020	CAMB-788CF	Pending	hCoV-19/England/CAMB-82103/2020	CAMB-82103
EPI_ISL_433788	hCoV-19/England/CAMB-788DE/2020	CAMB-788DE	Pending	hCoV-19/England/CAMB-82121/2020	CAMB-82121
EPI_ISL_433789	hCoV-19/England/CAMB-788ED/2020	CAMB-788ED	Pending	hCoV-19/England/CAMB-82130/2020	CAMB-82130
EPI_ISL_433790	hCoV-19/England/CAMB-788FC/2020	CAMB-788FC	Pending	hCoV-19/England/CAMB-8214F/2020	CAMB-8214F
EPI_ISL_433791	hCoV-19/England/CAMB-78917/2020	CAMB-78917	Pending	hCoV-19/England/CAMB-8216D/2020	CAMB-8216D
EPI_ISL_433792	hCoV-19/England/CAMB-78926/2020	CAMB-78926	Pending	hCoV-19/England/CAMB-8217C/2020	CAMB-8217C
EPI_ISL_433793	hCoV-19/England/CAMB-78935/2020	CAMB-78935	Pending	hCoV-19/England/CAMB-8218B/2020	CAMB-8218B
EPI_ISL_433794	hCoV-19/England/CAMB-78944/2020	CAMB-78944	Pending	hCoV-19/England/CAMB-821A9/2020	CAMB-821A9
EPI_ISL_433795	hCoV-19/England/CAMB-78953/2020	CAMB-78953	Pending	hCoV-19/England/CAMB-82188/2020	CAMB-82188
EPI_ISL_433796	hCoV-19/England/CAMB-78962/2020	CAMB-78962	Pending	hCoV-19/England/CAMB-821C7/2020	CAMB-821C7
EPI_ISL_433797	hCoV-19/England/CAMB-78971/2020	CAMB-78971	Pending	hCoV-19/England/CAMB-82200/2020	CAMB-82200
EPI_ISL_433798	hCoV-19/England/CAMB-78980/2020	CAMB-78980	Pending	hCoV-19/England/CAMB-8222E/2020	CAMB-8222E
EPI_ISL_433799	hCoV-19/England/CAMB-7899F/2020	CAMB-7899F	Pending	hCoV-19/England/CAMB-7540A/2020	CAMB-7540A
EPI_ISL_433800	hCoV-19/England/CAMB-789AE/2020	CAMB-789AE	Pending	hCoV-19/England/CAMB-756D7/2020	CAMB-756D7
EPI_ISL_433801	hCoV-19/England/CAMB-789BD/2020	CAMB-789BD	Pending	hCoV-19/England/CAMB-756F5/2020	CAMB-756F5
EPI_ISL_433802	hCoV-19/England/CAMB-789CC/2020	CAMB-789CC	Pending	hCoV-19/England/CAMB-75701/2020	CAMB-75701
EPI_ISL_433803	hCoV-19/England/CAMB-789DB/2020	CAMB-789DB	Pending	hCoV-19/England/CAMB-7572F/2020	CAMB-7572F
EPI_ISL_433804	hCoV-19/England/CAMB-789EA/2020	CAMB-789EA	Pending	hCoV-19/England/CAMB-7576B/2020	CAMB-7576B
EPI_ISL_433805	hCoV-19/England/CAMB-789F9/2020	CAMB-789F9	Pending	hCoV-19/England/CAMB-7577A/2020	CAMB-7577A
EPI_ISL_433806	hCoV-19/England/CAMB-78A05/2020	CAMB-78A05	Pending	hCoV-19/England/CAMB-75798/2020	CAMB-75798
EPI_ISL_433807	hCoV-19/England/CAMB-78A14/2020	CAMB-78A14	Pending	hCoV-19/England/CAMB-757A7/2020	CAMB-757A7
EPI_ISL_433808	hCoV-19/England/CAMB-78A32/2020	CAMB-78A32	Pending	hCoV-19/England/CAMB-757B6/2020	CAMB-757B6
EPI_ISL_433809	hCoV-19/England/CAMB-78A50/2020	CAMB-78A50	Pending	hCoV-19/England/CAMB-757C5/2020	CAMB-757C5
EPI_ISL_433810	hCoV-19/England/CAMB-78A7E/2020	CAMB-78A7E	Pending	hCoV-19/England/CAMB-757E3/2020	CAMB-757E3
EPI_ISL_433811	hCoV-19/England/CAMB-78A9C/2020	CAMB-78A9C	Pending	hCoV-19/England/CAMB-757F2/2020	CAMB-757F2
EPI_ISL_433812	hCoV-19/England/CAMB-78ABA/2020	CAMB-78ABA	Pending	hCoV-19/England/CAMB-7580E/2020	CAMB-7580E
EPI_ISL_433813	hCoV-19/England/CAMB-79A04/2020	CAMB-79A04	Pending	hCoV-19/England/CAMB-7582C/2020	CAMB-7582C
EPI_ISL_433814	hCoV-19/England/CAMB-79A22/2020	CAMB-79A22	Pending	hCoV-19/England/CAMB-7584A/2020	CAMB-7584A
EPI_ISL_433815	hCoV-19/England/CAMB-79A40/2020	CAMB-79A40	Pending	hCoV-19/England/CAMB-75859/2020	CAMB-75859
EPI_ISL_433816	hCoV-19/England/CAMB-79A5F/2020	CAMB-79A5F	Pending	hCoV-19/England/CAMB-75868/2020	CAMB-75868
EPI_ISL_433817	hCoV-19/England/CAMB-79A8C/2020	CAMB-79A8C	Pending	hCoV-19/England/CAMB-75886/2020	CAMB-75886
EPI_ISL_433818	hCoV-19/England/CAMB-79A9B/2020	CAMB-79A9B	Pending	hCoV-19/England/CAMB-75895/2020	CAMB-75895
EPI_ISL_433819	hCoV-19/England/CAMB-79AB9/2020	CAMB-79AB9	Pending	hCoV-19/England/CAMB-758A4/2020	CAMB-758A4
EPI_ISL_433820	hCoV-19/England/CAMB-79AC8/2020	CAMB-79AC8	Pending	hCoV-19/England/CAMB-758B3/2020	CAMB-758B3
EPI_ISL_433821	hCoV-19/England/CAMB-79AD7/2020	CAMB-79AD7	Pending	hCoV-19/England/CAMB-758C2/2020	CAMB-758C2
EPI_ISL_433822	hCoV-19/England/CAMB-79AE6/2020	CAMB-79AE6	Pending	hCoV-19/England/CAMB-758D1/2020	CAMB-758D1
EPI_ISL_433823	hCoV-19/England/CAMB-79AF5/2020	CAMB-79AF5	Pending	hCoV-19/England/CAMB-758E0/2020	CAMB-758E0
EPI_ISL_433824	hCoV-19/England/CAMB-79B01/2020	CAMB-79B01	Pending	hCoV-19/England/CAMB-758FF/2020	CAMB-758FF
EPI_ISL_433825	hCoV-19/England/CAMB-79B10/2020	CAMB-79B10	Pending	hCoV-19/England/CAMB-78522/2020	CAMB-78522
EPI_ISL_433826	hCoV-19/England/CAMB-7A8BE/2020	CAMB-7A8BE	Pending	hCoV-19/England/CAMB-78768/2020	CAMB-78768
EPI_ISL_433827	hCoV-19/England/CAMB-7A8DC/2020	CAMB-7A8DC	Pending	hCoV-19/England/CAMB-78759/2020	CAMB-78759
EPI_ISL_433828	hCoV-19/England/CAMB-7A8FA/2020	CAMB-7A8FA	Pending	hCoV-19/England/CAMB-79934/2020	CAMB-79934
EPI_ISL_433829	hCoV-19/England/CAMB-7AA30/2020	CAMB-7AA30	Pending	hCoV-19/England/CAMB-79943/2020	CAMB-79943
EPI_ISL_433830	hCoV-19/England/CAMB-7AA4F/2020	CAMB-7AA4F	Pending	hCoV-19/England/CAMB-79961/2020	CAMB-79961
EPI_ISL_433831	hCoV-19/England/CAMB-7AA6D/2020	CAMB-7AA6D	Pending	hCoV-19/England/CAMB-79970/2020	CAMB-79970
EPI_ISL_433832	hCoV-19/England/CAMB-7AA7C/2020	CAMB-7AA7C	Pending	hCoV-19/England/CAMB-7999E/2020	CAMB-7999E
EPI_ISL_433833	hCoV-19/England/CAMB-7AA8B/2020	CAMB-7AA8B	Pending	hCoV-19/England/CAMB-7999C/2020	CAMB-7999C
EPI_ISL_433834	hCoV-19/England/CAMB-7AAA9/2020	CAMB-7AAA9	Pending		
EPI_ISL_433835	hCoV-19/England/CAMB-7AAB8/2020	CAMB-7AAB8	Pending		
EPI_ISL_433836	hCoV-19/England/CAMB-7AAC7/2020	CAMB-7AAC7	Pending		
EPI_ISL_433837	hCoV-19/England/CAMB-7AAD6/2020	CAMB-7AAD6	Pending		
EPI_ISL_433838	hCoV-19/England/CAMB-7AAE5/2020	CAMB-7AAE5	Pending		
EPI_ISL_433839	hCoV-19/England/CAMB-7AAF4/2020	CAMB-7AAF4	Pending		
EPI_ISL_433840	hCoV-19/England/CAMB-7AB00/2020	CAMB-7AB00	Pending		
EPI_ISL_433841	hCoV-19/England/CAMB-7AB1F/2020	CAMB-7AB1F	Pending		
EPI_ISL_433842	hCoV-19/England/CAMB-7AB2E/2020	CAMB-7AB2E	Pending		
EPI_ISL_433843	hCoV-19/England/CAMB-7AB3D/2020	CAMB-7AB3D	Pending		
EPI_ISL_433844	hCoV-19/England/CAMB-7AB79/2020	CAMB-7AB79	Pending		
EPI_ISL_433845	hCoV-19/England/CAMB-7AB88/2020	CAMB-7AB88	Pending		
EPI_ISL_433846	hCoV-19/England/CAMB-7AB97/2020	CAMB-7AB97	Pending		
EPI_ISL_433847	hCoV-19/England/CAMB-7ABA6/2020	CAMB-7ABA6	Pending		
EPI_ISL_433848	hCoV-19/England/CAMB-7ABB5/2020	CAMB-7ABB5	Pending		
EPI_ISL_433849	hCoV-19/England/CAMB-7ABC4/2020	CAMB-7ABC4	Pending		
EPI_ISL_433850	hCoV-19/England/CAMB-7ABE2/2020	CAMB-7ABE2	Pending		
EPI_ISL_433851	hCoV-19/England/CAMB-7ABF1/2020	CAMB-7ABF1	Pending		
EPI_ISL_433852	hCoV-19/England/CAMB-7AC0D/2020	CAMB-7AC0D	Pending		
EPI_ISL_433853	hCoV-19/England/CAMB-7AC1C/2020	CAMB-7AC1C	Pending		
EPI_ISL_433854	hCoV-19/England/CAMB-7AC2B/2020	CAMB-7AC2B	Pending		