



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/193630/>

Version: Published Version

---

**Article:**

Colton, H., Parker, M.D., Stirrup, O. et al. (2023) Factors affecting turnaround time of SARS-CoV-2 sequencing for inpatient infection prevention and control decision making: analysis of data from the COG-UK HOCl study. *Journal of Hospital Infection*, 131. pp. 34-42. ISSN: 0195-6701

<https://doi.org/10.1016/j.jhin.2022.09.022>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# Factors affecting turnaround time of SARS-CoV-2 sequencing for inpatient infection prevention and control decision making: analysis of data from the COG-UK HOCl study

H. Colton<sup>a,b,\*</sup>, M.D. Parker<sup>c,d</sup>, O. Stirrup<sup>e</sup>, J. Blackstone<sup>f</sup>, M. Loose<sup>g</sup>, C.P. McClure<sup>g</sup>, S. Roy<sup>h</sup>, C. Williams<sup>h</sup>, J. McLeod<sup>i</sup>, D. Smith<sup>j</sup>, Y. Taha<sup>k</sup>, P. Zhang<sup>a</sup>, S.N. Hsu<sup>a,d</sup>, B. Kele<sup>l</sup>, K. Harris<sup>l</sup>, F. Mapp<sup>e</sup>, R. Williams<sup>h</sup>, COG-UK HOCl Investigators, COVID-19 Genomics UK (COG-UK) Consortium<sup>†</sup> P. Flowers<sup>i</sup>, J. Breuer<sup>h</sup>, D.G. Partridge<sup>b</sup>, T.I. de Silva<sup>a</sup>

<sup>a</sup> Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield, Sheffield, UK

<sup>b</sup> Directorate of Laboratory Medicine, Sheffield Teaching Hospitals NHS Foundation Trust, Sheffield, UK

<sup>c</sup> Sheffield Biomedical Research Centre, University of Sheffield, Sheffield, UK

<sup>d</sup> Sheffield Bioinformatics Core, University of Sheffield, Sheffield, UK

<sup>e</sup> Institute for Global Health, University College London, London, UK

<sup>f</sup> The Comprehensive Clinical Trials Unit, University College London, London, UK

<sup>g</sup> School of Life Sciences, University of Nottingham, Nottingham, UK

<sup>h</sup> Department of Infection, Immunity and Inflammation, Institute of Child Health, University College London, London, UK

<sup>i</sup> School of Psychological Sciences and Health, University of Strathclyde, Glasgow, UK

<sup>j</sup> Department of Applied Biology, Cellular and Molecular Sciences/Microbiology Group, Northumbria University, Newcastle, UK

<sup>k</sup> Department of Infection and Tropical Medicine, Newcastle Upon Tyne Hospitals NHS Foundation Trust, Royal Victoria Infirmary, Newcastle Upon Tyne, UK

<sup>l</sup> Virology Department, East and South East London Pathology Partnership, Royal London Hospital, Barts Health NHS Trust, London, UK

## ARTICLE INFO

### Article history:

Received 21 June 2022

Accepted 22 September 2022

Available online 10 October 2022

### Keywords:

Infection control  
Sequencing

## SUMMARY

**Background:** Barriers to rapid return of sequencing results can affect the utility of sequence data for infection prevention and control decisions.

**Aim:** To undertake a mixed-methods analysis to identify challenges that sites faced in achieving a rapid turnaround time (TAT) in the COVID-19 Genomics UK Hospital-Onset COVID-19 Infection (COG-UK HOCl) study.

**Methods:** For the quantitative analysis, timepoints relating to different stages of the sequencing process were extracted from both the COG-UK HOCl study dataset and surveys of study sites. Qualitative data relating to the barriers and facilitators to achieving rapid TATs were included from thematic analysis.

\* Corresponding author. Address: Department of Infection, Immunity and Cardiovascular Disease, Faculty of Medicine, Dentistry & Health, University of Sheffield, Medical School, Beech Hill Road, Sheffield S10 2RX, UK

E-mail address: [h.colton@sheffield.ac.uk](mailto:h.colton@sheffield.ac.uk) (H. Colton).

<sup>†</sup> Full list of consortium names and affiliations are in [Appendix 1](#) (see online supplementary material).

## SARS-CoV-2 Turnaround time



**Findings:** The overall TAT, from sample collection to receipt of sequence report by infection control teams, varied between sites (median 5.1 days, range 3.0–29.0 days). Most variation was seen between reporting of a positive COVID-19 polymerase chain reaction (PCR) result to sequence report generation (median 4.0 days, range 2.3–27.0 days). On deeper analysis, most of this variability was accounted for by differences in the delay between the COVID-19 PCR result and arrival of the sample at the sequencing laboratory (median 20.8 h, range 16.0–88.7 h). Qualitative analyses suggest that closer proximity of sequencing laboratories to diagnostic laboratories, increased staff flexibility and regular transport times facilitated a shorter TAT.

**Conclusion:** Integration of pathogen sequencing into diagnostic laboratories may help to improve sequencing TAT to allow sequence data to be of tangible value to infection control practice. Adding a quality control step upstream to increase capacity further down the workflow may also optimize TAT if lower quality samples are removed at an earlier stage.

© 2022 The Author(s). Published by Elsevier Ltd on behalf of The Healthcare Infection Society. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

## Introduction

The severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) pandemic has highlighted the utility of large-scale genomic sequencing to influence infection prevention and control (IPC) decisions [1,2]. While the technology to sequence pathogens rapidly using next-generation sequencing has been available for some time, its use has primarily been limited to genomic surveillance and retrospective transmission studies, often performed in large, centralised reference laboratories.

During the pandemic, the COVID-19 Genomics UK (COG-UK) Consortium established a network of sequencing hubs, pioneering a decentralised and distributed model of SARS-CoV-2 sequencing from National Health Service (NHS) hospitals [3]. The COG-UK Hospital-Onset COVID-19 Infection (COG-UK HOCl) study was nested within the COG-UK network, with the aim of assessing the impact of sequencing and its turnaround time (TAT) on several IPC outcomes [4]. The present authors recently reported that the likelihood of SARS-CoV-2 sequencing informing the IPC response to hospital-onset COVID-19 infection was dependent on the return of results within 5 days [5].

The time taken to return a potentially actionable sequence report to the IPC team is dependent on a variety of factors. For this paper, the authors further interrogated the data from the COG-UK HOCl study, alongside additional datapoints, in a post-hoc mixed-methods analysis, with the aim of identifying barriers to the achievement of a rapid sequencing TAT.

## Methods

### Background and design of the COG-UK HOCl study

The COG-UK HOCl study was a prospective non-randomised trial to evaluate the implementation and impact of SARS-CoV-2 sequencing on IPC practice. The study was approved by the National Research Ethics Service Committee – Cambridge South (REC 20/EE/0118) [4]. The study ran from December 2020 to April 2021 across 14 UK acute NHS hospital groups. The recruiting sites were linked to one of 11 sequencing laboratories where genomic sequencing took place.

The COG-UK HOCl study was split into baseline, rapid and longer turnaround phases to evaluate whether rapid

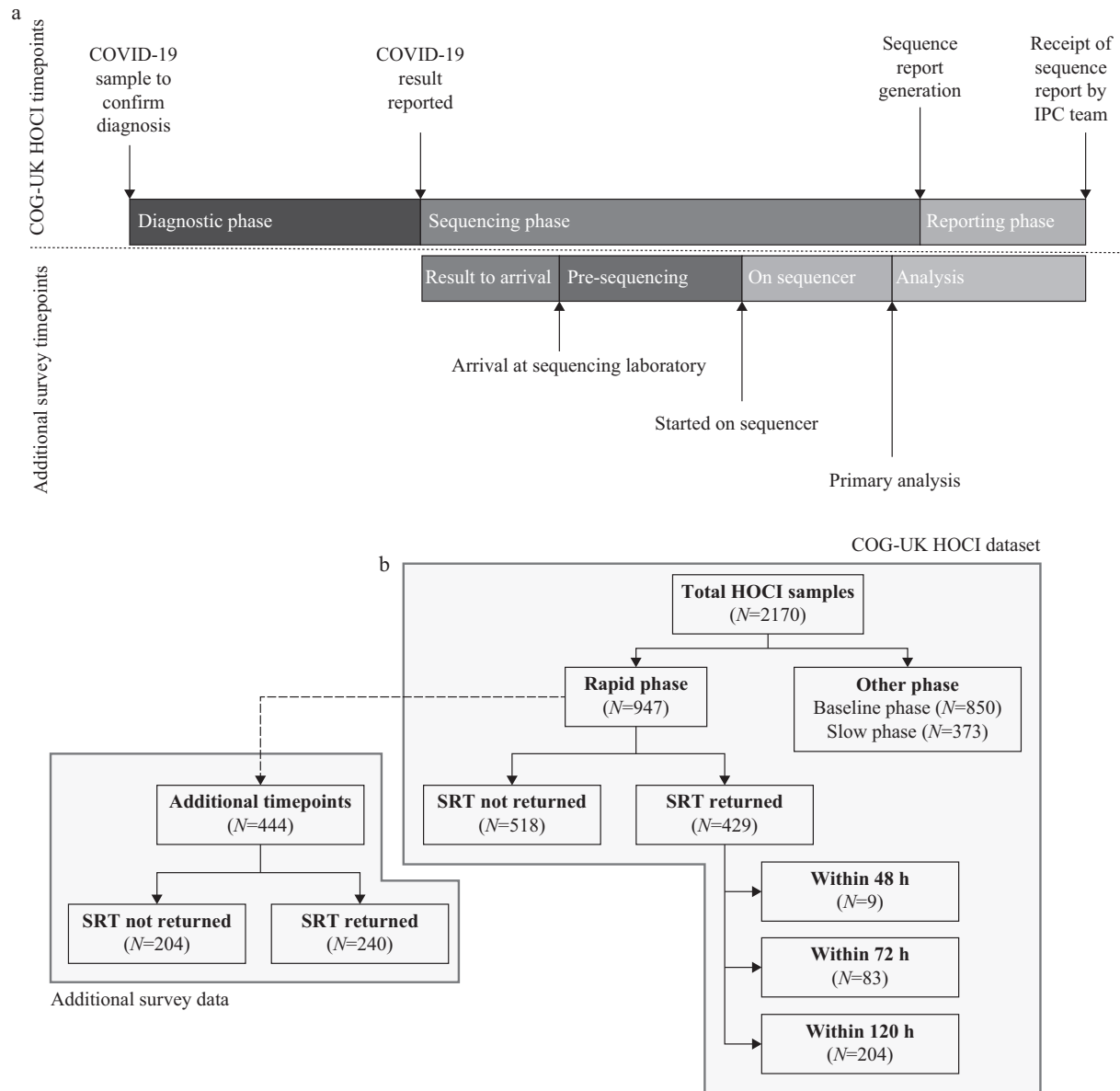
sequencing (i.e.  $\leq 48$  h) could improve IPC decision making in comparison with longer TAT (5–10 days), akin to using a centralised sequencing laboratory. Possible HOClIs were identified, and the respective samples were sent to the designated sequencing laboratory. A bespoke sequence report tool (SRT) was used to communicate the result to the IPC team for prospective action [6]. The SRT integrates genomic and epidemiological data from HOClIs to provide a one-page report identifying closely matched sequences within the hospital and at ward level, and assigns a probability estimate for nosocomial infection. Samples with genomic coverage  $< 90\%$  could not be used to generate an SRT [5,6].

Parallel independent quantitative and qualitative data collection and analysis were performed with subsequent integration of findings.

### Quantitative data extraction and analysis

For each sample, dates and times were extracted from the COG-UK HOCl study dataset for the following timepoints: (i) 'COVID-19 sample [taken] to confirm diagnosis'; (ii) 'COVID-19 result reported', (iii) 'Sequence report generation', and (iv) 'Receipt of sequence report by IPC team' (Figure 1a), in addition to patient study identifier, COG-UK ID, study site, and reason sequence was not returned within expected timeframes. Genomic coverage was extracted from the Cloud Infrastructure for Microbial Bioinformatics (CLIMB), and matched to each sample by COG-UK ID, where available.

Of the 2170 samples in the extract, only samples from the rapid phase of the COG-UK HOCl study were evaluated, when sites attempted to return an SRT within 48 h of sample collection ( $N=947$ , Figure 1b). Just under half of these samples had an SRT returned to the IPC team during the study ( $N=429/947$ , 45.3%). Reasons for failure to return an SRT included issues with sample quality/sequence failure ( $N=194/518$ , 37.5%), sequence report generation ( $N=130/518$ , 25.1%), processing delays ( $N=43/518$ , 8.3%), insufficient sample available ( $N=35/518$ , 6.8%) or unknown ( $N=116/518$ , 22.4%). For some timepoints, several samples only had the date available, with the time missing: 'COVID-19 sample [taken] to confirm diagnosis' ( $N=64/947$ , 6.8%), 'COVID-19 result reported' ( $N=18/947$ , 1.9%), 'Sequence report generation' ( $N=429/429$ , 100.0%) and

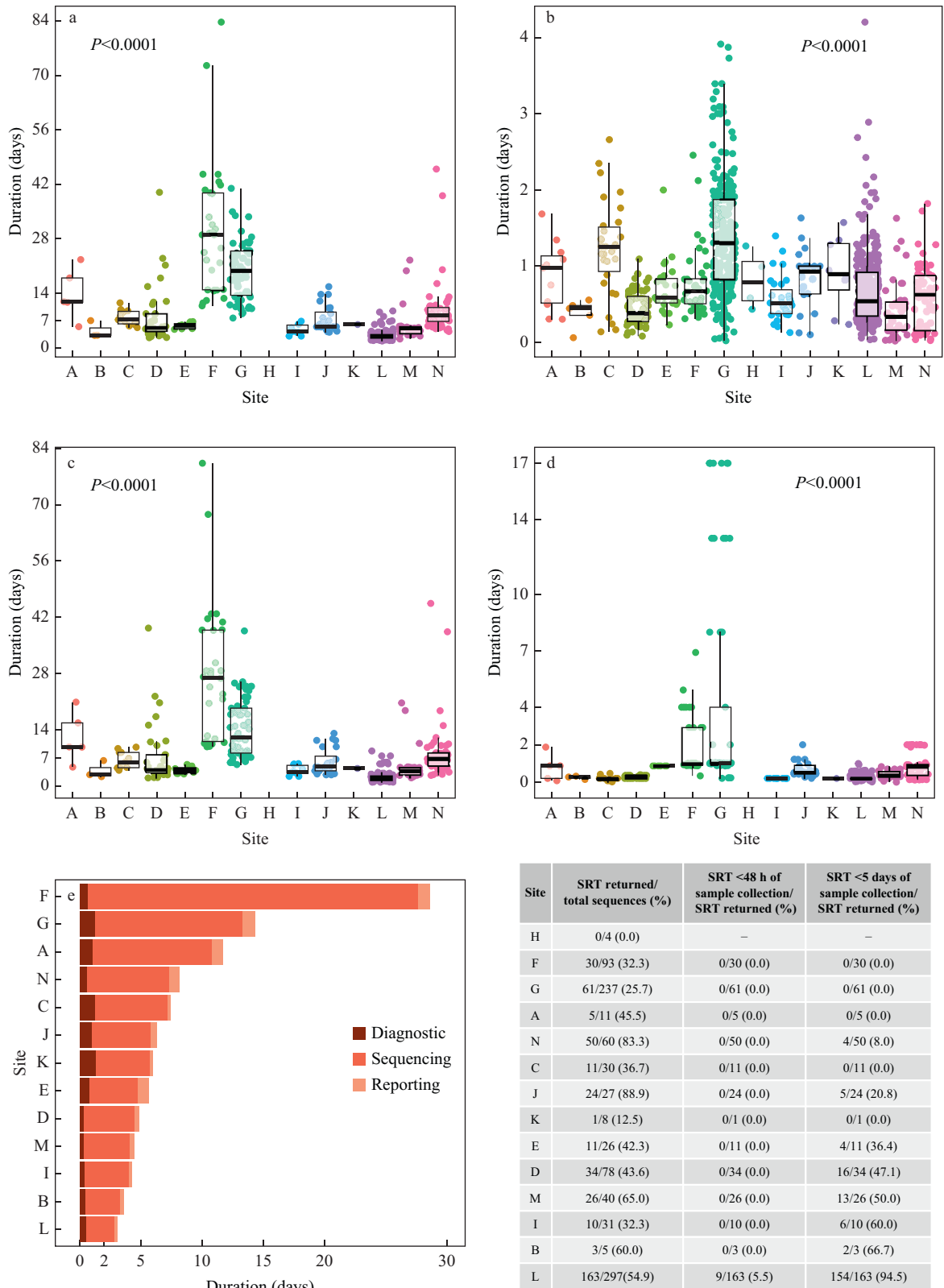


**Figure 1.** Summary of data used from the COVID-19 Genomics UK Hospital-Onset COVID-19 Infection (COG-UK HOCI) study dataset. (a) Phases of severe acute respiratory syndrome coronavirus-2 sequencing and available data from different timepoints. Pre-sequencing activity includes processes such as extraction (if required), amplicon generation for polymerase chain reaction, library preparation, and the time spent waiting in between each process. Primary analysis denotes processing of raw sequence data to generate a consensus sequence. (b) Breakdown of data acquired from main COG-UK HOCI study dataset and additional data acquired through survey of sites. COVID-19, coronavirus disease 2019; IPC, infection prevention and control; SRT, sequence report tool.

'Receipt of sequence report by IPC team' ( $N=146/429$ , 34.0%). Missing times for 'Sequence report generation' were estimated based on the 'Receipt of sequence report by IPC team' timepoint. If 'Receipt of sequence report by IPC team' was on the same date as 'Sequence report generation', corresponding missing 'Sequence report generation' times were replaced with either 00:00 if 'Receipt of sequence report by IPC team' was before 02:00 ( $N=4/429$ , 0.9%), or 07:00 if 'Receipt of sequence report by IPC team' was after 02:00 ( $N=288/429$ , 67.1%). All of the other missing times, including for the other timepoints, were replaced with 12:00 providing the dates were available, with sensitivity analyses undertaken to assess the impact of

missing data imputation (Figure S1, see online supplementary material). Samples where duration timepoints were unfeasible were either excluded from the analysis of their respective phase ( $N=67/947$ , 7.1%), or corrected where additional data were available from the site survey ( $N=7/947$ , 0.7%).

Additional data were requested from COG-UK HOCI study sites by e-mail invitation to complete a survey. Information requested included the type of sample received from the diagnostic laboratory (i.e. fresh unextracted or residual nucleic acid), frequency and method of transport between diagnostic and sequencing laboratories, number of sequencing runs per day, sequencing platform used, and availability



**Figure 2.** Scatterplots for all samples processed during the rapid phase using timepoints available within the COVID-19 Genomics UK Hospital-Onset COVID-19 Infection (COG-UK HOCl) study extract for (a) total turnaround time ( $N=429/429$ ), (b) diagnostic phase ( $N=880/947$ ), (c) sequencing phase ( $N=429/429$ ) and (d) reporting phase ( $N=429/429$ ). Boxplots represent medians plus interquartile ranges (IQR25 and IQR75). Sixty-seven samples were excluded from assessment of the diagnostic phase due to apparent errors in data entry. (e) Median durations for the diagnostic, sequencing and reporting phases by COG-UK HOCl site, using timepoints for the rapid phase samples in COG-UK HOCl extract which had a sequence report tool (SRT) returned ( $N=429/947$ ). The associated table shows the number of samples which were processed within the rapid phase per site.

of additional timepoints during the sequencing process (dates and times for when each sample arrived at the sequencing laboratory, when the sample was put on the sequencer, and when analysis of raw sequence data was commenced to generate a consensus SARS-CoV-2 sequence from each sample) (Figure 1a).

Seven of the 11 sequencing laboratories responded, and six agreed to provide additional timepoints where available. The sites that responded, and their median overall TATs, were Sites E (5.7 days), L (3.0 days), J (5.4 days), M (5.0 days), K (6.0 days), I (4.1 days) and A (11.9 days) (Table S1, see online supplementary material). These sequencing laboratories processed 444 of the COG-UK HOCl rapid phase samples ( $N=444/947$ , 46.9%), and just over half had an SRT returned ( $N=240/444$ , 54.1%) (Figure 1b). Reliable timepoints were available for 'Arrival at sequencing laboratory' and 'Time started on sequencer' for all six laboratories, which allowed deeper analysis of the sequencing phase from the COG-UK HOCl study dataset (Figure 1A). The sequencing laboratory for Sites E, J and K was only able to provide the dates for both of these timepoints, so times were estimated based on their standard practice. Although Site H responded to the survey, it was excluded from the analysis as it did not sequence any samples successfully within the rapid phase of the COG-UK HOCl study. If the duration between timepoints was illogical (i.e.  $<0$  h), the sample was excluded from analysis of that specific segment: PCR result to arrival at sequencing laboratory ( $N=3/444$ , 0.7%) and analysis ( $N=40/240$ , 16.7%).

Analysis of variance was used to calculate significance between sites for each of the durations in Figure 1a. Analysis was performed using R Studio (2021.09.1+372 'Ghost Orchid' Release).

### Qualitative design and analyses

Using a purposive subsample of five heterogeneous study sites, 39 diverse professional participants, all directly involved in implementing the COG-UK HOCl study, took part in semi-structured interviews between 23<sup>rd</sup> December 2020 and 2<sup>nd</sup> June 2021. Data collection focused on their HOCl experiences. A balance of deductive and inductive thematic analysis was conducted by a team of trained qualitative analysts. In this article, only the findings that relate to the barriers and facilitators to achieving rapid TATs are reported. The main results are presented by integrating the quantitative data on TATs, and qualitative findings where appropriate.

## Results

### Overall TAT

The overall TAT was calculated from 'COVID-19 sample [taken] to confirm diagnosis' to 'Receipt of sequence report by IPC team'. The TAT varied considerably between sites for the 429 samples which had an SRT returned to the IPC team (median 5.1 days, range 3.0–29.0 days;  $P<0.0001$ ) (Figure 2, Table S1, see online supplementary material). Sites self-reported if samples met their 'expected' TAT ( $N=209/429$ , 48.7%). However, the COG-UK HOCl study data revealed that only a fraction of SRTs were returned within 48 h of sample collection ( $N=9/429$ , 2.1%), with slightly more within

72 h ( $N=83/429$ , 19.3%) or 120 h of sample collection ( $N=204/429$ , 47.6%) (Figure 1b). The greatest intersite variation was seen within the sequencing phase (median 4.0 days, range 2.3–27.0 days;  $P<0.0001$ ), with less variation in the diagnostic (median 0.7 days, range 0.5–1.3 days;  $P<0.0001$ ) and reporting phases (median 0.4 days, range 0.2–1.0 days;  $P<0.0001$ ).

Qualitative analyses illuminated potential reasons for the low rates of meeting expected TATs, highlighting the fragility of the whole rapid SRT pipeline, where problems in any single step had consequences for others: 'you only need one thing to go wrong and it sort of snowballs really'. In this way, TAT was vulnerable to the effects of COVID-19, although automated processes and effective communication could help. Table 1 details qualitative barriers and facilitators to meeting rapid TATs in each phase. Many rate-limiting factors in the diagnostic phase related to the impacts of COVID-19 at the time of data collection, and reported teething troubles with diagnostic processes. Facilitators to the diagnostic phase focused on the efficient transporting of swabs to the diagnostic laboratory, and automated systems to pick up HOCl. For the reporting phase, barriers and facilitators related to both the generation and the dissemination of the report. Peer learning across sites generating reports and automated report generation were also notable facilitators.

### Detailed breakdown of the sequencing phase

The sequencing phase was broken down using the additional timepoints from the sites that responded to the survey, in order to further analyse sequencing laboratory activity (Figure 3).

#### COVID-19 result reported to arrival of sample at sequencing laboratory

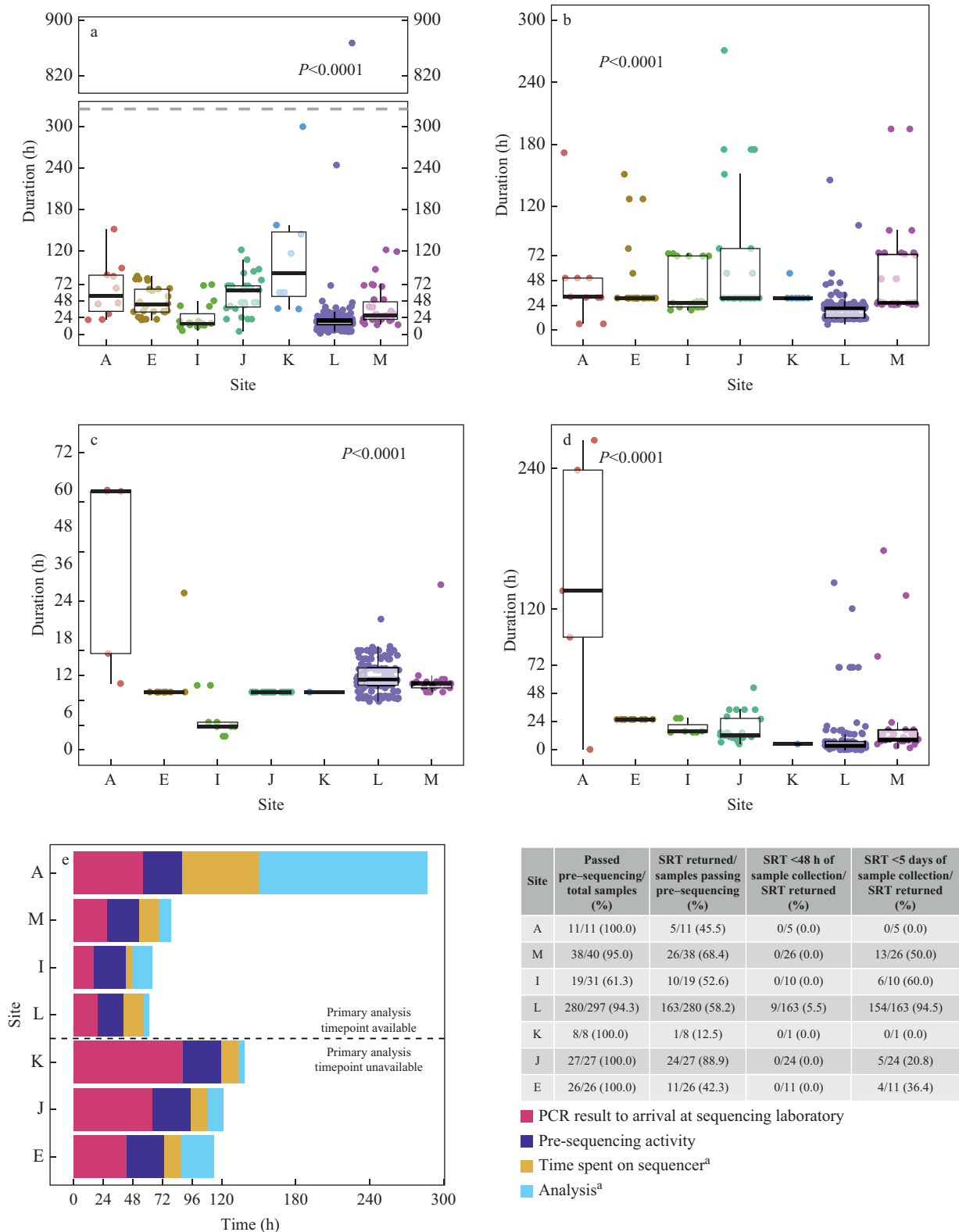
Large variability was seen between sites for the median duration between 'COVID-19 result reported' and 'Arrival at sequencing laboratory' (median 20.8 h, range 16.0–88.7 h;  $P<0.0001$ ; Table S2, see online supplementary material). The authors investigated whether the relative location of the diagnostic and sequencing laboratories could explain this variability. Site K was the furthest distance away from its sequencing laboratory (~205 km), and had the longest median time of 88.7 h. Sites I and L had the shortest median times of 16.0 h and 19.0 h, respectively, and their diagnostic and sequencing laboratories were much closer; in addition, the frequency of transport was greater at Site L (Table S3, see online supplementary material). Qualitative analyses suggest that the proximity of laboratories reduced delays, as did regular scheduled transport and dedicated pick-up times (Table 1). The ability of sequencing laboratory staff to work flexibly also enabled more rapid TATs.

#### Pre-sequencing

Pre-sequencing activity was calculated from 'Arrival at sequencing laboratory' through to 'Started on sequencer', and included extraction (if required), PCR, library preparation, and time between each process. All sequencing laboratories surveyed, except that for Site M, received fresh unextracted samples rather than residual nucleic acid. DNA quantification and normalisation was performed as a quality control (QC) step prior to library preparation at Sites L, M, I and A. Sites L, I and A performed DNA quantification on all samples, whilst Site M selected representative samples and their controls for testing.

**Table I**  
Barriers and facilitators to the main phases of rapid turnaround time (TAT)

Barriers	Facilitators	Indicative extract
<b>Overall turnaround time</b>		
<p>The volume and impact of COVID-19 infections stalled all steps and processes within TAT</p> <p>The number of steps within TAT meant if only one element was problematic, it had knock-on effects throughout</p>	<p>Flexibility of staff resource to be responsive to particular situational dynamics (ebbs and flows) reduced bottle necks in the multi-staged TAT process</p> <p>Close diagnostic and sequencing laboratories</p> <p>Robust workflow and automatization</p> <p>Skilled human resource</p> <p>Effective communication across the SRT pipeline</p>	<p>‘when you have 150 new cases a day, it just, it makes everything grind to a halt, you know, the patient flow in the hospital as well as specifically for the HOCl study, like even the diagnostic lab can’t cope, it has an impact on the research lab and the flow of samples. And then we have so many samples, the sequencing isn’t as good, means more samples failing and you know, handling the data’s a lot harder’</p>
<b>Diagnostic phase</b>		
<p>COVID-19 demands made it difficult to collect samples (e.g. volume of patients)</p> <p>Volume of samples meant delays to testing (competing testing streams) in diagnostic laboratories</p> <p>Staff meeting called for each potential HOCl (at start of trial) delayed classification</p> <p>Untrained staff appraising clinical notes without expert knowledge made classification difficult (at start of trial)</p> <p>Reliance on a clinician for classification slowed down process (at start of trial)</p>	<p>Having porters available to transport swabs to diagnostic laboratory</p> <p>Regular transport system between sites and laboratories (e.g. taxis)</p> <p>Automated systems to pick up HOClS were developed once screening established</p> <p>Established IPC team at hospital made the classifications</p>	<p>‘It felt was a bit like a brick wall a lot of the time unfortunately. Just the systems and the way it works and the fact that we take swabs and they have to be couriered over to [hospital name B]. It takes a bit longer over here; we don’t quite get the turnaround’</p>
<b>Sequencing phase</b>		
<p>Staff absences made it difficult to get samples to sequencing laboratory</p> <p>Volume of HOCl samples to sequence and the prioritization of healthcare worker samples make it difficult to generate sequence data file</p> <p>Samples going missing because of test validation studies made it difficult to generate the sequence data file</p> <p>Samples with high CT values wasted sequence capacity</p> <p>Volume of sequencing occurring limited sequencing</p>	<p>Dedicated pick-up time from diagnostic laboratory meant sufficient time at sequencing laboratory</p> <p>Having sequencing laboratory close by reduced delays</p> <p>High-quality, committed sequencing laboratory teams were able to offer flexible working patterns</p> <p>Dedicated resource assisted with sequencing phase</p>	<p>‘they worked really late that night to make sure that things were still kept on track. So I think that’s, I think that’s an amazing effort on their part’</p>
<b>Reporting phase</b>		
<p>IT problems with memory and grid lines</p> <p>Problems with updating tools for lineage</p> <p>The tool was not designed for the amount of data</p> <p>Access to GLUE delayed report generation</p> <p>There were problems sending the PDF of the report to some e-mail accounts (NHS particularly)</p>	<p>Input from bioinformatician and IT</p> <p>Communication and peer-learning across sites running reports</p> <p>Automated report generation and dissemination – to password protected HOCl slack channel/teams (workplace messaging systems) and necessary NHS emails</p>	<p>‘I think it’s a GLUE server that’s been down and over the weekend apparently the sys[tem] admin don’t kind of, the CLIMB admin don’t work, so we were kind of stuck where we had worked quite hard to get sequence data out and we can actually get the reporting tool out .... And that’s a bit frustrating I guess when something out of your hands goes wrong and you’ve done everything possible to try and, yeah’</p>



**Figure 3.** Durations within the sequencing phase for each sample by site for (a) polymerase chain reaction (PCR) result to arrival at sequencing laboratory ( $N=406/444$ ), (b) pre-sequencing ( $N=406/444$ ), (c) time spent on sequencer<sup>a</sup> ( $N=239/240$ ), and (d) sequence analysis<sup>a</sup> ( $N=200/240$ ). Boxplots represent interquartile range (IQR)25, median and IQR75. For 3a, the y axis was broken in order to show outliers using the R package ggbreak [13]. (e) Median durations and number of samples for each stage of process from 'COVID-19 result reported' onwards for the samples processed within the rapid phase from the surveyed sites ( $N=444/947$ ). As the 'Primary analysis' timepoint was not available for Sites E, J and K, the 'Sequence report generation' timepoint from the COVID-19 Genomics UK Hospital-Onset COVID-19 Infection study dataset was used in lieu, which corresponds to Figure 3c–e <sup>a</sup>For time spent on sequencer and analysis, only samples which had a sequence report tool (SRT) returned were used ( $N=240/444$ ), as unsuccessful sequences may artificially shorten the duration if the run was aborted.

Sites I and A performed additional QC steps. Site I screened cycle threshold (CT) values from the diagnostic PCR, and based on their experience, only proceeded if CT was <35. Site A checked PCR amplicons on gel prior to DNA quantification, and only processed them if a visible band was present. The median pre-sequencing activity was relatively consistent between sites (22.3 h, range 21.2–32.0 h;  $P < 0.0001$ ) (Table S2, see online supplementary material). Accurate start times for extraction, PCR and library preparation were unavailable; however, Sites L, M and A self-reported times based on typical practice (Table S4, see online supplementary material). Site L was the only site that reported performing two library preparations per day. Only minimal differences were seen between the estimated library preparation times by sequencing platform (Table S2, see online supplementary material).

#### On sequencer

The time spent on the sequencer was calculated from the 'Started on sequencer' timepoint through to the 'Primary analysis' timepoint, as the time the sequencing run was stopped was not readily available. The 'Primary analysis' timepoint was not available for Sites E, J and K, so the 'Sequence report generation' timepoint was used in lieu. The median time spent on the sequencer for samples where the SRT was returned was 17.1 h (range 5.7–62.6 h;  $P < 0.0001$ ) (Table S2, see online supplementary material). Site I had the shortest median time on the sequencer, and reported starting sequencing their samples around 12:00–16:00 to start primary analysis the same evening. Site A had the longest median time spent on the sequencer, and reported having too many other samples to process, or having issues with the reporting tools CLIMB and GLUE ( $N = 9/11$ , 81.8%) [7].

#### Analysis

The duration of analysis was calculated from 'Primary analysis' (or 'Sequence report generation' for Sites E, J and K as mentioned previously) through to 'Receipt of sequence report by IPC team'. The median duration of analysis across all sites was 4.6 h, with a range of 2.9–135.8 h ( $P < 0.0001$ ; Table S2, see online supplementary material). The sites that reported using DNA quantification as a QC step prior to library preparation had higher percentages of SRT return within 5 days than sites that did not have a QC step (Figure 3E).

## Discussion

The COG-UK HOCl study found that returning an SRT within 5 days changed the actions of the IPC teams in approximately 20% of HOClIs [5]. This mixed-methods analysis found that many sites did not manage to return any of their SRTs within 5 days, and identified some of the challenges that sites faced.

As the greatest intersite variability was seen in the time between the diagnostic PCR result and the arrival of the sample at the sequencing laboratory, an obvious factor to optimise TAT would be to reduce the distance and/or increase the transport frequency between diagnostic and sequencing laboratories, as described in the qualitative data. Integrating sequencing into diagnostic laboratories could be an ideal solution, and would also facilitate the transfer of patient level data including current location and prior ward movements from patient administration systems, and provide the geotemporal data required for easy and rapid interpretation of sequence reports.

Integrated laboratories have also been reported to increase regional and national processing power for the surveillance of antimicrobial resistance [8]. Where integration is not possible, reducing the distance between laboratories and regular dedicated transport times would allow laboratories to plan their workflows in order to optimise TAT and reduce the likelihood of missing samples.

The second greatest variability was seen in the median duration between the start of primary analysis and receipt of the SRT by the IPC team. As sites reported they were overwhelmed with processing other samples during the COVID-19 pandemic, adding a QC step may increase capacity further down the workflow; however, caution should be applied if CT values are used, given that significant variability between laboratories has been reported [9]. Additionally, if sites were able to run samples on the sequencer earlier in the day, the sequencing process could be stopped and its output analysed within the same day, allowing a shorter sequencing and analysis time.

Outside of the COVID-19 response, there have been successful reports of the use of rapid sequencing to influence the IPC response for other pathogens of interest, such as methicillin-resistant *Staphylococcus aureus*, *Clostridioides difficile* and vancomycin-resistant enterococci [10,11]. For each pathogen, laboratories would have to consider pathogen-specific factors which could affect TAT, such as the time required for culturing bacteria or the frequency of performing other tests, such as immunoassays. The pressure on the microbiology laboratory workload would also vary at different points of the year, and outside of a pandemic, would likely also have an impact on TAT.

This study was limited by missing and erroneous times within the datasets; however, the authors were able to mitigate this, in part, through either correcting, estimating or excluding timepoints. In addition, the potential of volunteer bias within the survey data is recognised, as sites with shorter TATs were more likely to respond and provide additional data. Although the authors were unable to conclude whether genomic coverage was affected by either the sample type received by the sequencing laboratory, or the time between sample collection and extraction, it is well reported that RNA is at risk of degradation if samples are not processed promptly [12].

In conclusion, IPC interventions in response to presumed nosocomial transmission events are often resource intensive in terms of human, financial and operational impact, and thus practice developments which confirm or refute case linkage within a clinically meaningful time scale have the potential to be of great benefit to healthcare services.

These results present evidence supporting the integration of pathogen sequencing into diagnostic laboratories, in order for sequence data to be of tangible value to IPC practice.

Laboratories using rapid sequencing for IPC purposes may be able to utilise these findings to streamline and optimise their own workflows for SARS-CoV-2 and other pathogens. The challenges and optimal TATs of integrated sequencing for IPC use on a larger scale need further analysis for other pathogens, including whether challenges faced by sites would be similar outside of a pandemic if a short TAT is desired.

## Acknowledgements

The authors wish to thank the Principal Investigators of the COG-UK HOCl study: Dr Anu Chawla, Dr Maria-Teresa Cutino-

Moguel, Prof Alison Holmes, Dr Nick Machin, Dr Nikunj Mahida, Dr Taibtha Mahungu, Dr Gaia Nebbia, Dr Cassie Pope, Dr Kordo Saeed, Dr Tranprit Saluja, Dr Gee Yen Shin, Dr Thushan de Silva, Dr Yusri Taha, Prof Emma Thomson and their teams for their time and support. The authors also acknowledge the support of the independent members of the Joint Trial Steering Committee and Data Monitoring Committee (Prof Marion Koopmans, Prof Walter Zingg, Prof Colm Bergin, Prof Karla Hemming, Prof Katherine Fielding), as well as the non-independent members (Prof Nick Lemoine, Prof Sharon Peacock). The authors also wish to thank members of COG-UK who supported the COG-UK HOCl study directly: Dr Ewan Harrison, Dr Katerina Galai, Dr Frances Coll, Dr Michael Chapman, Prof Thomas Connor, Prof Nick Loman and their teams. Finally, the authors wish to thank the COG-UK Consortium and the NIHR Clinical Research Network.

#### Conflict of interest statement

Since the completion of this work, MP has become an employee of Oxford Nanopore Technologies plc.

#### Funding sources

COG-UK is supported by funding from the Medical Research Council part of UK Research & Innovation, the National Institute of Health Research (Grant MC\_PC\_19027), and Genome Research Limited, operating as the Wellcome Sanger Institute. The authors acknowledge the use of data generated through the COVID-19 Genomics Programme funded by the Department of Health and Social Care. The views expressed are those of the authors, and not necessarily those of the Department of Health and Social Care or the UK Health Security Agency.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhin.2022.09.022>.

## References

- [1] Walker A, Houwaart T, Finzer P, Ehlkes L, Tyshaieva A, Damagnez M, et al. Characterization of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection clusters based on integrated genomic surveillance, outbreak analysis and contact tracing in an urban setting. *Clin Infect Dis* 2022;74:1039–46.
- [2] Francis RV, Billam H, Clarke M, Yates C, Tsoferidis T, Berry L, et al. The impact of real-time whole-genome sequencing in controlling healthcare-associated SARS-CoV-2 outbreaks. *J Infect Dis* 2022;225:10–8.
- [3] Covid-19 Genomics UK (COG-UK) Consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* 2020;1:e99–100.
- [4] Blackstone J, Stirrup O, Mapp F, Panca M, Copas A, Flowers P, et al. Protocol for the COG-UK hospital-onset COVID-19 infection (HOCl) multicentre interventional clinical study: evaluating the efficacy of rapid genome sequencing of SARS-CoV-2 in limiting the spread of COVID-19 in UK NHS hospitals. *BMJ Open* 2022;12:e052514.
- [5] Stirrup O, Blackstone J, Mapp F, MacNeil A, Panca M, Holmes A, et al. Effectiveness of rapid SARS-CoV-2 genome sequencing in supporting infection control for hospital-onset COVID-19 infection: multicentre, prospective study *eLife* 2022;11:e78427.
- [6] Stirrup OT, Hughes J, Parker M, Partridge DG, Shepherd JG, Blackstone J, et al. Rapid feedback on hospital onset SARS-CoV-2 infections combining epidemiological and sequencing data. *Elife* 2021;10:1–30.
- [7] Singer JB, Thomson EC, McLauchlan J, Hughes J, Gifford RJ. GLUE: a flexible software system for virus sequence data. *BMC Bioinformatics* 2018;19:1–18.
- [8] Kekre M, Arevalo SA, Valencia MF, Lagrada ML, Macaranas PKV, Nagaraj G, et al. Integrating scalable genome sequencing into microbiology laboratories for routine antimicrobial resistance surveillance. *Clin Infect Dis* 2021;73:S258–66.
- [9] Evans D, Cowen S, Kammel M, O’Sullivan DM, Stewart G, Grunert HP, et al. The dangers of using Cq to quantify nucleic acid in biological samples: a lesson from COVID-19. *Clin Chem* 2021;68:153–62.
- [10] Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. *BMJ Open* 2012;2:1–9.
- [11] McGann P, Bunin JL, Snesrud E, Singh S, Maybank R, Ong AC, et al. Real time application of whole genome sequencing for outbreak investigation – what is an achievable turnaround time? *Diagn Microbiol Infect Dis* 2016;85:277–82.
- [12] Lewandowski K, Bell A, Miles R, Carne S, Wooldridge D, Manso C, et al. The effect of nucleic acid extraction platforms and sample storage on the integrity of viral RNA for use in whole genome sequencing. *J Mol Diagnostics* 2017;19:303–12.
- [13] Xu Shuangbin, Chen Meijun, Feng Tingze, Zhan Li, Zhou Lang YG. Use ggbreak to effectively utilize plotting space to deal with large datasets and outliers. *Front Genet* 2021;12. <https://doi.org/10.3389/fgene.2021.774846>.