

Early analysis of a potential link between viral load and the N501Y mutation in the SARS-COV-2 spike protein

Tanya Golubchik^{*1}, Katrina A. Lythgoe^{*1}, Matthew Hall¹, Luca Ferretti¹, Helen R. Fryer¹, George MacIntyre-Cockett^{1,2}, Mariateresa de Cesare^{1,2}, Amy Trebes², Paolo Piazza², David Buck², John A. Todd², The COVID-19 Genomics UK (COG-UK) consortium³, Christophe Fraser¹, David Bonsall^{1,2}

*Equal contribution

Corresponding authors: David.Bonsall@bdi.ox.ac.uk, Tanya.Golubchik@bdi.ox.ac.uk

¹Big Data Institute, Nuffield Department of Medicine, University of Oxford, Old Road Campus, Oxford OX3 7FL, UK.

²Wellcome Centre for Human Genetics, Nuffield Department of Medicine, NIHR Biomedical Research Centre, University of Oxford, Old Road Campus, Oxford OX3 7BN, UK.

³www.cogconsortium.uk. Full list of names and affiliations are in the Appendix.

Abstract

A new variant of SARS-CoV-2 has emerged which is increasing in frequency, primarily in the South East of England (lineage B.1.1.7 (1); VUI-202012/01). One potential hypothesis is that infection with the new variant results in higher viral loads, which in turn may make the virus more transmissible. We found higher (sequence derived) viral loads in samples from individuals infected with the new variant with median inferred viral loads were three-fold higher in individuals with the new variant. Most of the new variants were sampled in Kent and Greater London. We observed higher viral loads in Kent compared to Greater London for both the new variant and other circulating lineages. Outside Greater London, the variant has higher viral loads, whereas within Greater London, the new variant does not have significantly higher viral loads compared to other circulating lineages. Higher variant viral loads outside Greater London could be due to demographic effects, such as a faster variant growth rate compared to other lineages or concentration in particular age-groups. However, our analysis does not exclude a causal link between infection with the new variant and higher viral loads. This is a preliminary analysis and further work is needed to investigate any potential causal link between infection with this new variant and higher viral loads, and whether this results in higher transmissibility, severity of infection, or affects relative rates of symptomatic and asymptomatic infection

Document Description and Purpose:

This is an updated report submitted to NERVTAG in December 2020 as part of urgent investigations into the new variant of SARS-CoV-2 (VUI-202012/01). It makes full use of (and is restricted to) all sequence data and associated metadata available to us at the time this original report was submitted and remains provisional. Under normal circumstances more genomes and metadata would be obtained and included before making this report public. We will update this preprint when more genomes and metadata are available and before submitting for peer review.

Background

On 14 December 2020 a new variant of SARS-CoV-2 circulating in the UK was reported (2, 3), characterised by the N501Y mutation in the receptor binding domain (RBD) of Spike, the Δ H69/V70 deletion, and numerous other mutations (1). The rise in frequency of this variant is associated with a sharp increase in reported cases in the South East of England, raising concerns that the variant could be more transmissible. We performed a rapid analysis to investigate whether the new variant is associated with higher viral loads, since higher viral loads may indicate increased transmissibility.

Methods

As members of the COG-UK consortium (<https://www.cogconsortium.uk/>), we sequenced RT-QPCR SARS-CoV-2 positive samples originating from four UK Lighthouse laboratories, which provide Pillar 2 COVID-19 testing services. The samples were sequenced using veSEQ, our quantitative sequencing approach for which the number of unique mapped reads is correlated with, and thus can be used as a proxy for, viral load. For a full description of the sequencing protocol see (4, 5).

We used \log_{10} (mapped reads) as a proxy for viral load (see fig S1 in (5)). Comparisons between distributions of \log_{10} (mapped reads) were made using Welsh t-test (two-tailed), with p-values combined using Stouffer's method where appropriate. We also performed a multivariate logistic regression analysis.

Number of unique mapped reads is negatively correlated with Ct value

Given the known negative correlation between viral load and cycle threshold (Ct) values (6) obtained during PCR testing (7), we first confirmed a strong negative correlation between \log_{10} (unique mapped reads) and Ct values for samples that we sequenced from Lighthouse laboratories (linear regression, $r^2=0.43$, $p<0.001$, Fig. 1).

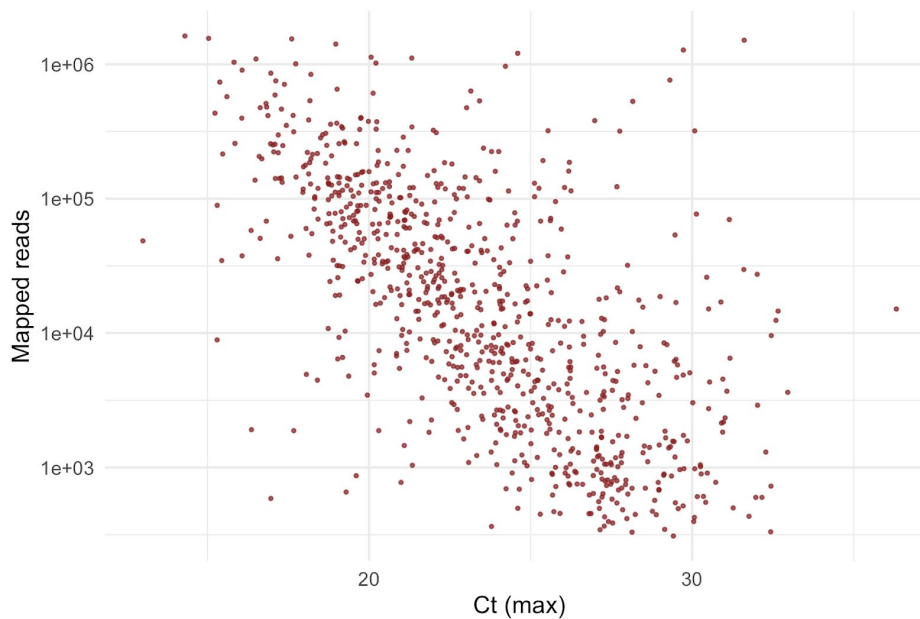


Figure 1. A strong negative correlation between Ct value and \log_{10} (number of mapped reads). Number of uniquely mapped reads per sample can be used as a proxy for viral load. The Ct value shown is the maximum Ct value obtained from Majora (the COG database) from retrospective data for all Lighthouse laboratories that supply Ct data; \log_{10} of uniquely mapped (deduplicated) reads obtained with veSEQ platform correlates well with Ct. This does not include samples in this report since Ct values were not yet available.

The new variant is associated with higher viral loads

The N501Y mutation is strongly linked with other mutations characterising the new variant (VUI-202012/01) in our dataset, including the Δ H69/V70 deletion, and therefore we used Y501 as a marker of the new variant. The Δ H69/V70 deletion alone is not a specific marker of VUI-202012/01 in our data, while lineage B.1.1.70, which is currently present in Wales and in some cases carries Y501 but never the deletion, was not present in our data.

We identified 88 samples that produced consensus sequences with the Y501 variant. All variant samples were taken between 31 Oct 2020 and 13 Nov 2020, and therefore we only considered samples (Y501 and N501) taken during this period, since Ct values have been shown to vary by calendar time (7).

When comparing the number of unique mapped reads in the Y501 variant samples (median \log_{10} (reads)=4.64, $N=88$) with that in the to N501 samples (median \log_{10} (reads)=4.16, $N=1299$), we found higher counts in the former (Welch t -Test $p=0.014$; Fig. 1). This is equivalent to around 3-fold higher median viral loads in the Y501 variant samples compared to N501 samples.

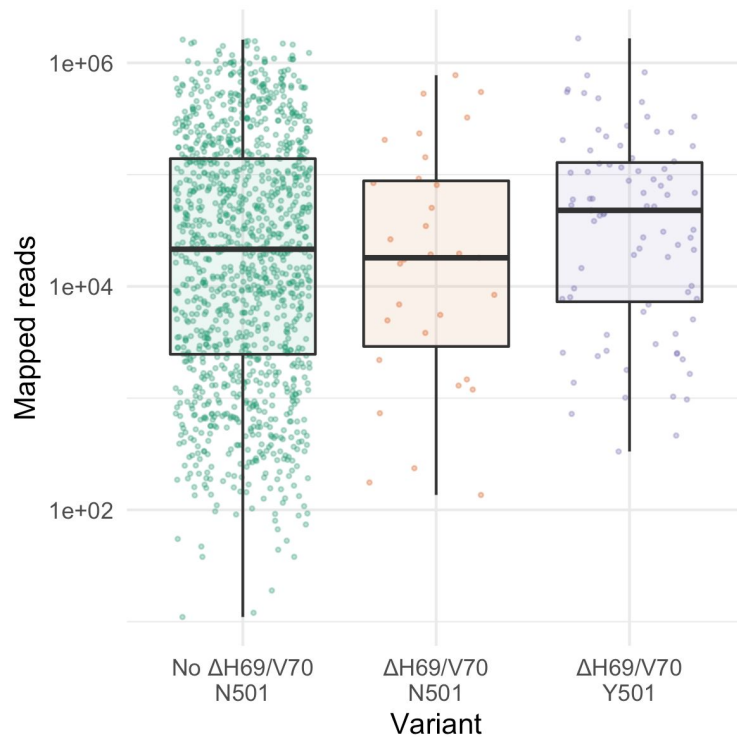


Figure 1: Higher numbers of mapped reads in samples with the Y501 variant. Box and scatter plots of unique mapped reads, stratified according to variant. Points within each batch are jittered to aid visualisation. Horizontal lines in boxplots represent the median and the interquartile range. The Y501 variant has a higher number of mapped reads, whereas the Δ H69/V70 deletion only has a higher number of mapped reads in the presence of the N501Y mutation.

This result remained significant when we controlled for batch (Fig. 3a, $p=0.011$, combined p -value via Stouffer's method), but not Lighthouse laboratory (Fig. 3b, $p=0.052$). The correlation between the new variant and viral load is also associated with a relative paucity of samples with lower ($<10^3$) mapped reads among Y501 samples (Fig. 1, $p=0.0053$, chi-squared test; 10^3 logged mapped reads is equivalent to a viral load of $\sim 10^4$ copies per reaction, max Ct ~ 28). When comparing samples with just the Δ H69/V70 deletion (without the Y501 variant) to samples without the deletion, we did not find a significant difference in $\log_{10}(\text{reads})$ ($p=0.86$; controlling for batch $p=0.56$, and for Lighthouse lab $p=0.54$) (Fig. 1).

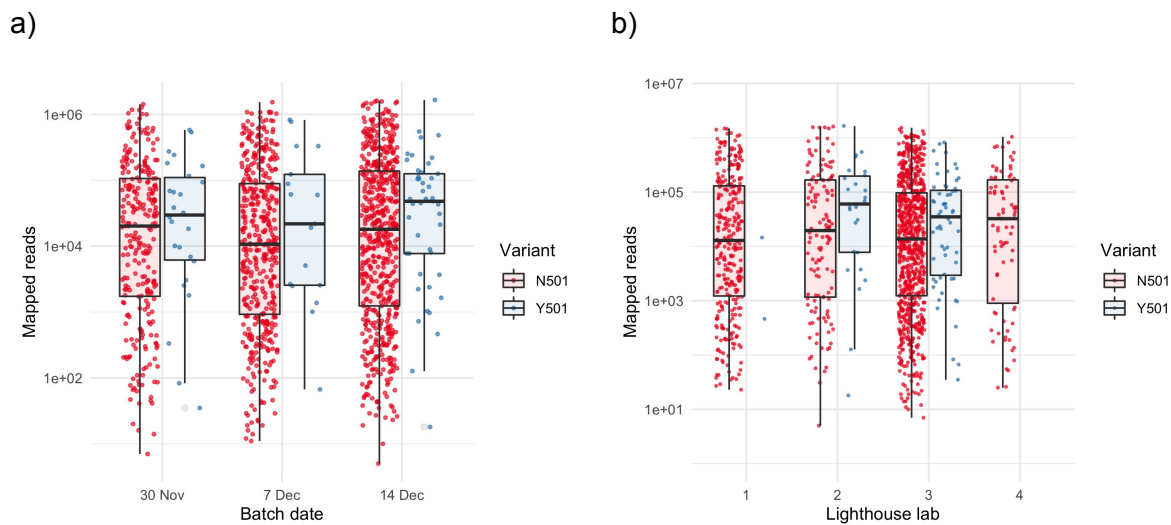


Figure 3. Higher numbers of mapped reads in samples exhibiting the Y501 variant

Box and scatter plots of unique mapped reads, stratified by (a) batch date and (b) anonymised Lighthouse lab. There is no significant difference among batches or Lighthouse labs for N501 samples ($p > 0.1$ for all pairwise comparisons). Points within each batch are jittered to aid visualisation. Horizontal lines in boxplots represent the median and the interquartile range.

Viral loads differ by sampling location

To test whether the difference in viral loads for samples with the new variant could in part be explained by geographic effects, we considered the sampling location (adm2 district) where this information was available. Of the 88 Y501 variants sampled, 24 were in Greater London, 46 in Kent, and in lower numbers ($N=1-5$) in other areas (Bristol, Essex, Hampshire, Leicestershire, Norfolk, Surrey and West Sussex). Regardless of variant presence, all samples from Greater London had significantly lower viral loads than those from other locations ($p=0.0016$, Welch's t -test), and the association between Y501 and higher viral load was not significant in this region ($p=0.91$; Fig. 4). Outside Greater London, viral loads for Y501 were significantly higher than for N501 ($p=0.0068$). Within Kent, the location with the greatest number of Y501 samples, Y501 viral loads were not significantly higher than N501 viral loads ($p=0.089$). These results indicate a correlation between infection with the new variant (VUI-202012/01) and (inferred) viral load outside Greater London, although we are currently underpowered to draw firm conclusions. The lack of association within Greater London could be due to lack of power, or to demographic or epidemiological differences in London compared with the other locations.

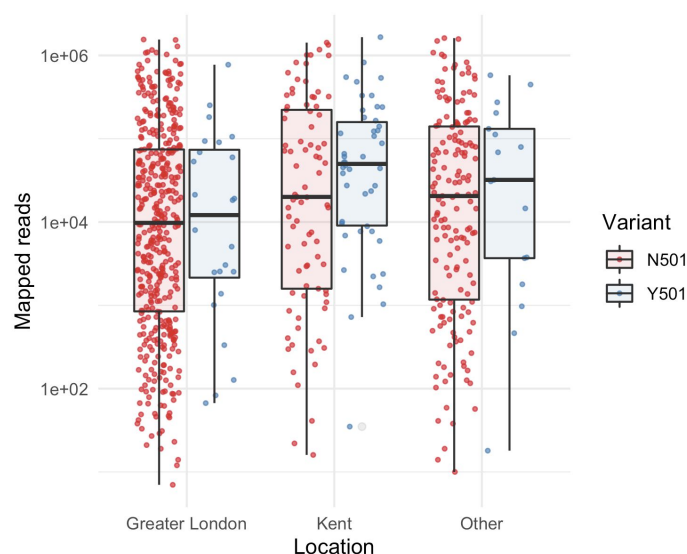


Figure 4. Number of mapped reads varies by sampling location

Box and scatter plots of unique mapped reads, stratified by sampling location. Points within each batch are jittered to aid visualisation. Horizontal lines in boxplots represent the median and the interquartile range. Only sampling locations with at least one Y501 sample were included.

In a multivariate logistic regression analysis for variables associated with higher viral load (Table 1), the Y501 variant was associated with a fivefold increase in odds of $\geq 10^3$ mapped reads ($p=0.036$). The fitted model with interaction terms suggest a much smaller effect of the variant outside Kent, with the total odds increase reduced to 1.75 for Greater London and 1.24 for other regions, but the interaction term coefficients were not statistically significant ($p=0.27$ and $p=0.16$, respectively). Thus, if the association of the variant with a paucity of low viral load samples is stronger in Kent compared to other areas (e.g. due to epidemiological, demographic, or sampling differences), we lack the necessary power to demonstrate it. No other variables showed evidence of an association.

Table 1. Logistic regression analysis, identifying variables associated with $\geq 10^3$ mapped reads.

	Odds ratio (95% CI)
Y105	5.078* (1.111, 23.218)
Region (reference: Kent)	
Greater London	0.675 (0.371, 1.227)
Other	0.869 (0.476, 1.588)
Lighthouse lab (reference: 1)	
2	1.148 (0.703, 1.874)
3	1.183 (0.610, 2.296)
4	0.906 (0.512, 1.605)
Batch date (reference: 30 Nov)	
7 Dec	0.822 (0.575, 1.173)
14 Dec	0.955 (0.502, 1.816)
Interaction: Y105 × Greater London	0.345 (0.053, 2.248)
Interaction: Y105 × Other region	0.245 (0.034, 1.771)
Observations	1,376
Note:	*p<0.05; **p<0.01; ***p<0.001

Odds ratios for each variable and 95% confidence intervals for those ratios are presented. Lighthouse labs are anonymised as in Fig. 2.

Caveats and Limitations

This is a preliminary analysis, and other factors could explain the (inferred) higher viral loads in samples with the new variant (VUI-202012/01), in addition to a working hypothesis that there is a causal effect of the new variant on within-host virus abundance. Whether the correlation is causative (infections with the new variant have higher viral loads) or correlative (e.g. due to epidemiological dynamics, demographics of individuals infected with the new variant, and/or sampling) warrants further study.

Individuals contributing samples in this analysis were tested as part of the test and trace program, which is primarily aimed towards individuals seeking a test following the onset of symptoms. We observed a broad spectrum of viral loads among the samples we sequenced. Given known associations between lower viral loads and later infection (8), and higher viral loads at the onset of symptoms, this suggests our full dataset consists of individuals in both early and late stages of symptomatic infection. Whilst we do not *a priori* expect there to be a systematic difference in the timing of sampling relative to infection, in an exponentially growing population the expectation is to sample relatively more people early in infection (9). Whether or not early sampling-bias supports an effect on inferred viral loads will depend on the relative epidemiological dynamics of the new and other variants. If, for example, VUI-202012/01 is growing faster, this could result in a bias for it to be sampled earlier. This is consistent with the relative paucity of VUI-202012/01 samples with low viral load.

In addition, VUI-202012/01 might be circulating within particular demographics (e.g. age groups) that tend to have higher viral loads when sampled. This may explain the apparently different patterns in Greater London and elsewhere. Focussed transmission within a particular demographic group is also more likely during the early stages of epidemic growth of a given lineage, before it disperses into the wider population. We were unable to test these hypotheses as we did not have demographic data relating to the sampled individuals with the new variant. We also cannot rule out other additional confounding effects and recommend that such effects are investigated further.

Future prospects

A number of processes could have caused the rapid growth of the new variant (VUI-202012/01), including founder effects, or biological mechanisms that increase its transmissibility. Higher viral loads are one such potential mechanism: Transmissibility of viruses is understood to be higher in individuals who exhibit higher viral loads (10) and in HIV viral load is partly determined by virus genotype (11). Our observation of higher inferred viral loads in individuals infected with the new variant suggests that increased transmissibility of the new variant is plausible, but important caveats remain.

We recommend further investigations to evaluate this hypothesis. We note that we have used Y501 as a marker for the new variant; a large number of other mutations also characterise this new variant lineage (1), and therefore Y501 *per se* might not be causing the effect (if there is one). We also note that higher viral loads can be associated with higher levels of viral virulence, and therefore links between the new variant and the severity of infection should be monitored carefully (12).

Whether or not observed higher viral loads associated with this variant are a direct cause of infection with the variant, a consequence of faster epidemic growth, or linked to particular demographics, our data are consistent with rapid growth of this specific lineage.

References

1. A. Rambaut *et al.*, Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations (2020), (available at <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>).
2. S. A. Kemp, R. P. Datir, D. A. Collier, I. Ferreira, A. Carabelli, W. Harvey, D. L. Robertson, R. K. Gupta, Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion Δ H69/V70. *Cold Spring Harbor Laboratory* (2020), p. 2020.12.14.422555.
3. Update on new SARS-CoV-2 variant and how COG-UK tracks emerging mutations - COG-UK Consortium, (available at https://www.cogconsortium.uk/news_item/update-on-new-sars-cov-2-variant-and-how-cog-uk-tracks-emerging-mutations/).
4. D. Bonsall, T. Golubchik, M. de Cesare, M. Limbada, B. Kosloff, G. MacIntyre-Cockett, M. Hall, C. Wymant, M. Azim Ansari, L. Abeler-Dörner, A. Schaap, A. Brown, E. Barnes, E. Piwowar-Manning, S. Eshleman, E. Wilson, L. Emel, R. Hayes, S. Fidler, H. Ayles, R. Bowden, C. Fraser, HPTN 071 (PopART) Team, A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. *J. Clin. Microbiol.* **58** (2020), doi:10.1128/JCM.00382-20.
5. K. A. Lythgoe, M. Hall, L. Ferretti, M. de Cesare, G. MacIntyre-Cockett, A. Trebes, M. Andersson, N. Otecko, E. L. Wise, N. Moore, J. Lynch, S. Kidd, N. Cortes, M. Mori, R. Williams, G. Vernet, A. Justice, A. Green, S. M. Nicholls, M. Azim Ansari, L. Abeler-Dörner, C. E. Moore, T. E. A. Peto, D. W. Eyre, R. Shaw, P. Simmonds, D. Buck, J. A. Todd, on behalf of OVSG Analysis Group, T. R. Connor, A. da Silva Filipe, J. Shepherd, E. C. Thomson, The COVID-19 Genomics UK (COG-UK) consortium, D. Bonsall, C. Fraser, T. Golubchik, Within-host genomics of SARS-CoV-2. *Cold Spring*

Harbor Laboratory (2020), p. 2020.05.28.118992.

6. Public Health England, Cycle threshold (Ct) in SARS-CoV-2 RT-PCR (2020), (available at <https://www.gov.uk/government/publications/cycle-threshold-ct-in-sars-cov-2-rt-pcr>).
7. A. Sarah Walker, E. Pritchard, T. House, J. V. Robotham, P. J. Birrell, I. Bell, J. I. Bell, J. N. Newton, J. Farrar, I. Diamond, R. Studley, J. Hay, K.-D. Vihta, T. Peto, N. Stoesser, P. C. Matthews, D. W. Eyre, K. B. Pouwels, the COVID-19 Infection Survey team, Viral load in community SARS-CoV-2 cases varies widely and temporally. *medRxiv*, 2020.10.25.20219048 (2020).
8. S. M. Kissler, J. R. Fauver, C. Mack, S. W. Olesen, C. Tai, K. Y. Shiue, C. C. Kalinich, S. Jednak, I. M. Ott, C. B. F. Vogels, J. Wohlgemuth, J. Weisberger, J. DiFiori, D. J. Anderson, J. Mancell, D. D. Ho, N. D. Grubaugh, Y. H. Grad, SARS-CoV-2 viral dynamics in acute infections. *medRxiv*, 2020.10.21.20217042 (2020).
9. J. A. Hay, L. Kennedy-Shaffer, S. Kanjilal, M. Lipsitch, M. J. Mina, Estimating epidemiologic dynamics from single cross-sectional viral load distributions. *medRxiv*, 2020.10.08.20204222 (2020).
10. X. He, E. H. Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao, Y. C. Lau, J. Y. Wong, Y. Guan, X. Tan, X. Mo, Y. Chen, B. Liao, W. Chen, F. Hu, Q. Zhang, M. Zhong, Y. Wu, L. Zhao, F. Zhang, B. J. Cowling, F. Li, G. M. Leung, Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **26**, 672–675 (2020).
11. C. Fraser, K. Lythgoe, G. E. Leventhal, G. Shirreff, T. Déirdre Hollingsworth, S. Alizon, S. Bonhoeffer, Virulence and Pathogenesis of HIV-1 Infection: An Evolutionary Perspective. *Science*. **343** (2014), doi:10.1126/science.1243727.
12. Public Health England, PHE investigating a novel strain of COVID-19. *GOV.UK* (2020), (available at <https://www.gov.uk/government/news/phe-investigating-a-novel-strain-of-covid-19>).

Acknowledgments

We are grateful to Lorne Lornie, Angie Green, The Oxford Genomics Centre and The Wellcome Centre for Human Genetics, for all their support in generating the data for this study.

Ethics

The COVID-19 Genomics UK (COG-UK) consortium study protocol was approved by the Public Health England Research Ethics and Governance Group (reference: R&D NR0195) on the 8th of April 2020. The decision was communicated to Professor Sharon Peacock by Dr Elizabeth Coates, Head of Research Governance, PHE Research Support and Governance Office.

Funding

The UK COVID-19 Genomics 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. The research was supported by the Wellcome Trust Core Award Grant Number 203141/Z/16/Z with funding from the NIHR Oxford BRC. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department

of Health. KAL and HRF are supported by The Wellcome Trust and The Royal Society (107652/Z/15/Z) and by Li Ka Shing Foundation funding awarded to KL. TG, MH, LF, MdC, GMC, CF and DB are supported by Li Ka Shing Foundation funding awarded to CF.

Appendix: COG-UK Full list of consortium names and affiliations

Funding acquisition, leadership, supervision, metadata curation, project administration, samples, logistics, Sequencing, analysis, and Software and analysis tools:

Thomas R Connor^{33,34}, and Nicholas J Loman¹⁵.

Leadership, supervision, sequencing, analysis, funding acquisition, metadata curation, project administration, samples, logistics, and visualisation:

Samuel C Robson⁶⁸.

Leadership, supervision, project administration, visualisation, samples, logistics, metadata curation and software and analysis tools:

Tanya Golubchik²⁷.

Leadership, supervision, metadata curation, project administration, samples, logistics sequencing and analysis:

M. Estee Torok^{8,10}.

Project administration, metadata curation, samples, logistics, sequencing, analysis, and software and analysis tools:

William L Hamilton^{8,10}.

Leadership, supervision, samples logistics, project administration, funding acquisition sequencing and analysis:

David Bonsall²⁷.

Leadership and supervision, sequencing, analysis, funding acquisition, visualisation and software and analysis tools:

Ali R Awan⁷⁴.

Leadership and supervision, funding acquisition, sequencing, analysis, metadata curation, samples and logistics:

Sally Corden³³.

Leadership supervision, sequencing analysis, samples, logistics, and metadata curation:

Ian Goodfellow¹¹.

Leadership, supervision, sequencing, analysis, samples, logistics, and Project administration:

Darren L Smith ^{60, 61} .

Project administration, metadata curation, samples, logistics, sequencing and analysis:

Martin D Curran ¹⁴ , and Surendra Parmar ¹⁴ .

Samples, logistics, metadata curation, project administration sequencing and analysis:

James G Shepherd ²¹ .

Sequencing, analysis, project administration, metadata curation and software and analysis tools:

Matthew D Parker ³⁸ and Dinesh Aggarwal ^{1, 2, 3} .

Leadership, supervision, funding acquisition, samples, logistics, and metadata curation:

Catherine Moore ³³ .

Leadership, supervision, metadata curation, samples, logistics, sequencing and analysis:

Derek J Fairley ^{6, 88} , Matthew W Loose ⁵⁴ , and Joanne Watkins ³³ .

Metadata curation, sequencing, analysis, leadership, supervision and software and analysis tools:

Matthew Bull ³³ , and Sam Nicholls ¹⁵ .

Leadership, supervision, visualisation, sequencing, analysis and software and analysis tools:

David M Aanensen ^{1, 30} .

Sequencing, analysis, samples, logistics, metadata curation, and visualisation:

Sharon Glaysher ⁷⁰ .

Metadata curation, sequencing, analysis, visualisation, software and analysis tools:

Matthew Bashton ⁶⁰ , and Nicole Pacchiarini ³³ .

Sequencing, analysis, visualisation, metadata curation, and software and analysis tools:

Anthony P Underwood ^{1, 30} .

Funding acquisition, leadership, supervision and project administration:

Thushan I de Silva ³⁸ , and Dennis Wang ³⁸ .

Project administration, samples, logistics, leadership and supervision:

Monique Andersson²⁸, Anoop J Chauhan⁷⁰, Mariateresa de Cesare²⁶, Catherine Ludden^{1,3}, and Tabitha W Mahungu⁹¹.

Sequencing, analysis, project administration and metadata curation:

Rebecca Dewar²⁰, and Martin P McHugh²⁰.

Samples, logistics, metadata curation and project administration:

Natasha G Jesudason²¹, Kathy K Li MBBCh²¹, Rajiv N Shah²¹, and Yusri Taha⁶⁶.

Leadership, supervision, funding acquisition and metadata curation:

Kate E Templeton²⁰.

Leadership, supervision, funding acquisition, sequencing and analysis:

Simon Cottrell³³, Justin O'Grady⁵¹, Andrew Rambaut¹⁹, and Colin P Smith⁹³.

Leadership, supervision, metadata curation, sequencing and analysis:

Matthew T.G. Holden⁸⁷, and Emma C Thomson²¹.

Leadership, supervision, samples, logistics and metadata curation:

Samuel Moses^{81,82}.

Sequencing, analysis, leadership, supervision, samples and logistics:

Meera Chand⁷, Chrystala Constantinidou⁷¹, Alistair C Darby⁴⁶, Julian A Hiscox⁴⁶, Steve Paterson⁴⁶, and Meera Unnikrishnan⁷¹.

Sequencing, analysis, leadership and supervision and software and analysis tools:

Andrew J Page⁵¹, and Erik M Volz⁹⁶.

Samples, logistics, sequencing, analysis and metadata curation:

Charlotte J Houldcroft⁸, Aminu S Jahun¹¹, James P McKenna⁸⁸, Luke W Meredith¹¹, Andrew Nelson⁶¹, Sarojini Pandey⁷², and Gregory R Young⁶⁰.

Sequencing, analysis, metadata curation, and software and analysis tools:

Anna Price³⁴, Sara Rey³³, Sunando Roy⁴¹, Ben Temperton⁴⁹, and Matthew Wyles³⁸.

Sequencing, analysis, metadata curation and visualisation:

Stefan Rooke¹⁹, and Sharif Shaaban⁸⁷.

Visualisation, sequencing, analysis and software and analysis tools:

Helen Adams³⁵, Yann Bourgeois⁶⁹, Katie F Loveson⁶⁸, Áine O'Toole¹⁹, and Richard Stark⁷¹.

Project administration, leadership and supervision:

Ewan M Harrison ^{1,3}, David Heyburn ³³, and Sharon J Peacock ^{2,3}

Project administration and funding acquisition:

David Buck ²⁶, and Michaela John³⁶

Sequencing, analysis and project administration:

Dorota Jamrozy ¹, and Joshua Quick ¹⁵

Samples, logistics, and project administration:

Rahul Batra ⁷⁸, Katherine L Bellis ^{1,3}, Beth Blane ³, Sophia T Girgis ³, Angie Green ²⁶, Anita Justice ²⁸, Mark Kristiansen ⁴¹, and Rachel J Williams ⁴¹.

Project administration, software and analysis tools:

Radoslaw Poplawski¹⁵.

Project administration and visualisation:

Garry P Scarlett ⁶⁹.

Leadership, supervision, and funding acquisition:

John A Todd ²⁶, Christophe Fraser ²⁷, Judith Breuer ^{40,41}, Sergi Castellano ⁴¹, Stephen L Michell ⁴⁹, Dimitris Gramatopoulos ⁷³, and Jonathan Edgeworth ⁷⁸.

Leadership, supervision and metadata curation:

Gemma L Kay ⁵¹.

Leadership, supervision, sequencing and analysis:

Ana da Silva Filipe ²¹, Aaron R Jeffries ⁴⁹, Sascha Ott ⁷¹, Oliver Pybus ²⁴, David L Robertson ²¹, David A Simpson ⁶, and Chris Williams ³³.

Samples, logistics, leadership and supervision:

Cressida Auckland ⁵⁰, John Boyes ⁸³, Samir Dervisevic ⁵², Sian Ellard ^{49,50}, Sonia Goncalves¹, Emma J Meader ⁵¹, Peter Muir ², Husam Osman ⁹⁵, Reenesh Prakash ⁵², Venkat Sivaprakasam ¹⁸, and Ian B Vipond ².

Leadership, supervision and visualisation

Jane AH Masoli ^{49,50}.

Sequencing, analysis and metadata curation

Nabil-Fareed Alikhan⁵¹, Matthew Carlile⁵⁴, Noel Craine³³, Sam T Haldenby⁴⁶, Nadine Holmes⁵⁴, Ronan A Lyons³⁷, Christopher Moore⁵⁴, Malorie Perry³³, Ben Warne⁸⁰, and Thomas Williams¹⁹.

Samples, logistics and metadata curation:

Lisa Berry⁷², Andrew Bosworth⁹⁵, Julianne Rose Brown⁴⁰, Sharon Campbell⁶⁷, Anna Casey¹⁷, Gemma Clark⁵⁶, Jennifer Collins⁶⁶, Alison Cox^{43, 44}, Thomas Davis⁸⁴, Gary Eltringham⁶⁶, Cariad Evans^{38,39}, Clive Graham⁶⁴, Fenella Halstead¹⁸, Kathryn Ann Harris⁴⁰, Christopher Holmes⁵⁸, Stephanie Hutchings², Miren Iturriza-Gomara⁴⁶, Kate Johnson^{38,39}, Katie Jones⁷², Alexander J Keeley³⁸, Bridget A Knight^{49,50}, Cherian Koshy⁹⁰, Steven Liggett⁶³, Hannah Lowe⁸¹, Anita O Lucaci⁴⁶, Jessica Lynch^{25,29}, Patrick C McClure⁵⁵, Nathan Moore³¹, Matilde Mori^{25,29,32}, David G Partridge^{38,39}, Pinglawathee Madona^{43,44}, Hannah M Pymont², Paul Anthony Randell^{43,44}, Mohammad Raza^{38,39}, Felicity Ryan⁸¹, Robert Shaw²⁸, Tim J Sloan⁵⁷, and Emma Swindells⁶⁵.

Sequencing, analysis, Samples and logistics:

Alexander Adams³³, Hibo Asad³³, Alec Birchley³³, Tony Thomas Brooks⁴¹, Giselda Bucca⁹³, Ethan Butcher⁷⁰, Sarah L Caddy¹³, Laura G Caller^{2,3,12}, Yasmin Chaudhry¹¹, Jason Coombes³³, Michelle Cronin³³, Patricia L Dyal⁴¹, Johnathan M Evans³³, Laia Fina³³, Bree Gatica-Wilcox³³, Iliana Georgana¹¹, Lauren Gilbert³³, Lee Graham³³, Danielle C Groves³⁸, Grant Hall¹¹, Ember Hilvers³³, Myra Hosmillo¹¹, Hannah Jones³³, Sophie Jones³³, Fahad A Khokhar¹³, Sara Kumziene-Summerhayes³³, George MacIntyre-Cockett²⁶, Rocio T Martinez Nunez⁹⁴, Caoimhe McKerr³³, Claire McMurray¹⁵, Richard Myers⁷, Yasmin Nicole Panchbhaya⁴¹, Malte L Pinckert¹¹, Amy Plimmer³³, Joanne Stockton¹⁵, Sarah Taylor³³, Alicia Thornton⁷, Amy Trebes²⁶, Alexander J Trotter⁵¹, Helena Jane Tutill⁴¹, Charlotte A Williams⁴¹, Anna Yakovleva¹¹ and Wen C Yew⁶².

Sequencing, analysis and software and analysis tools:

Mohammad T Alam⁷¹, Laura Baxter⁷¹, Olivia Boyd⁹⁶, Fabricia F. Nascimento⁹⁶, Timothy M Freeman³⁸, Lily Geidelberg⁹⁶, Joseph Hughes²¹, David Jorgensen⁹⁶, Benjamin B Lindsey³⁸, Richard J Orton²¹, Manon Ragonnet-Cronin⁹⁶ Joel Southgate^{33,34}, and Sreenu Vattipally²¹.

Samples, logistics and software and analysis tools:

Igor Starinskij²³.

Visualisation and software and analysis tools:

Joshua B Singer²¹, Khalil Abudahab^{1,30}, Leonardo de Oliveira Martins⁵¹, Thanh Le-Viet⁵¹, Mirko Menegazzo³⁰, Ben EW Taylor^{1,30}, and Corin A Yeats³⁰.

Project Administration:

Sophie Palmer³, Carol M Churcher³, Alisha Davies³³, Elen De Lacy³³, Fatima Downing³³, Sue Edwards³³, Nikki Smith³⁸, Francesc Coll⁹⁷, Nazreen F Hadjirin³ and Frances Bolt^{44,45}.

Leadership and supervision:

Alex Alderton¹, Matt Berriman¹, Ian G Charles⁵¹, Nicholas Cortes³¹, Tanya Curran⁸⁸, John Danesh¹, Sahar Eldirdiri⁸⁴, Ngozi Elumogo⁵², Andrew Hattersley^{49,50}, Alison Holmes^{44,45}, Robin Howe³³, Rachel Jones³³, Anita Kenyon⁸⁴, Robert A Kingsley⁵¹, Dominic Kwiatkowski^{1,9}, Cordelia Langford¹, Jenifer Mason⁴⁸, Alison E Mather⁵¹, Lizzie Meadows⁵¹, Sian Morgan³⁶, James Price^{44,45}, Trevor I Robinson⁴⁸, Giri Shankar³³, John Wain⁵¹, and Mark A Webber⁵¹.

Metadata curation:

Declan T Bradley^{5,6}, Michael R Chapman^{1,3,4}, Derrick Crooke²⁸, David Eyre²⁸, Martyn Guest³⁴, Huw Gulliver³⁴, Sarah Hoosdally²⁸, Christine Kitchen³⁴, Ian Merrick³⁴, Siddharth Mookerjee^{44,45}, Robert Munn³⁴, Timothy Peto²⁸, Will Potter⁵², Dheeraj K Sethi⁵², Wendy Smith⁵⁶, Luke B Snell^{75,94}, Rachael Stanley⁵², Claire Stuart⁵² and Elizabeth Wastenge²⁰.

Sequencing and analysis:

Erwan Acheson⁶, Safiah Afifi³⁶, Elias Allara^{2,3}, Roberto Amato¹, Adrienn Angyal³⁸, Elihu Aranday-Cortes²¹, Cristina Ariani¹, Jordan Ashworth¹⁹, Stephen Attwood²⁴, Alp Aydin⁵¹, David J Baker⁵¹, Carlos E Balcazar¹⁹, Angela Beckett⁶⁸, Robert Beer³⁶, Gilberto Betancor⁷⁶, Emma Betteridge¹, David Bibby⁷, Daniel Bradshaw⁷, Catherine Bresner³⁴, Hannah E Bridgewater⁷¹, Alice Broos²¹, Rebecca Brown³⁸, Paul E Brown⁷¹, Kirstyn Bruncker²², Stephen N Carmichael²¹, Jeffrey K. J. Cheng⁷¹, Dr Rachel Colquhoun¹⁹, Gavin Dabrera⁷, Johnny Debebe⁵⁴, Eleanor Drury¹, Louis du Plessis²⁴, Richard Eccles⁴⁶, Nicholas Ellaby⁷, Audrey Farbos⁴⁹, Ben Farr¹, Jacqueline Findlay⁴¹, Chloe L Fisher⁷⁴, Leysa Marie Forrest⁴¹, Sarah Francois²⁴, Lucy R. Frost⁷¹, William Fuller³⁴, Eileen Gallagher⁷, Michael D Gallagher¹⁹, Matthew Gemmell⁴⁶, Rachel AJ Gilroy⁵¹, Scott Goodwin¹, Luke R Green³⁸, Richard Gregory⁴⁶, Natalie Groves⁷, James W Harrison⁴⁹, Hassan Hartman⁷, Andrew R Hesketh⁹³, Verity Hill¹⁹, Jonathan Hubb⁷, Margaret Hughes⁴⁶, David K Jackson¹, Ben Jackson¹⁹, Keith James¹, Natasha Johnson²¹, Ian Johnston¹, Jon-Paul Keatley¹, Moritz Kraemer²⁴, Angie Lackenby⁷, Mara Lawniczak¹, David Lee⁷, Rich Livett¹, Stephanie Lo¹, Daniel Mair²¹, Joshua Maksimovic³⁶, Nikos Manesis⁷, Robin Manley⁴⁹, Carmen Manso⁷, Angela Marchbank³⁴, Inigo Martincorena¹, Tamyo Mbisa⁷, Kathryn McCluggage³⁶, JT McCrone¹⁹, Shahjahan Miah⁷, Michelle L Michelsen⁴⁹, Mari Morgan³³, Gaia Nebbia⁷⁸, Charlotte Nelson⁴⁶, Jenna Nichols²¹, Paola Niola⁴¹, Kyriaki Nomikou²¹, Steve Palmer¹, Naomi Park¹, Yasmin A Parr¹, Paul J Parsons³⁸, Vineet Patel⁷, Minal Patel¹, Clare Pearson^{2,1}, Steven Platt⁷, Christoph Puethe¹, Mike Quail¹, Jayna Raghvani²⁴, Lucille Rainbow⁴⁶, Shavanthi Rajatileka¹, Mary Ramsay⁷, Paola C Resende Silva^{41,42}, Steven Rudder⁵¹, Chris Ruis³, Christine M Sambles⁴⁹, Fei Sang⁵⁴, Ulf Schaefer⁷, Emily Scher¹⁹, Carol Scott¹, Lesley Shirley¹, Adrian W Signell⁷⁶, John Sillitoe¹, Christen Smith¹, Dr Katherine L Smollett²¹, Karla Spellman³⁶, Thomas D Stanton¹⁹, David J Studholme⁴⁹, Grace Taylor-Joyce⁷¹, Ana P Tedim⁵¹, Thomas Thompson⁶,

Nicholas M Thomson⁵¹, Scott Thurston¹, Lily Tong²¹, Gerry Tonkin-Hill¹, Rachel M Tucker³⁸, Edith E Vamos⁴, Tetyana Vasylyeva²⁴, Joanna Warwick-Dugdale⁴⁹, Danni Weldon¹, Mark Whitehead⁴⁶, David Williams⁷, Kathleen A Williamson¹⁹, Harry D Wilson⁷⁶, Trudy Workman³⁴, Muhammad Yasir⁵¹, Xiaoyu Yu¹⁹, and Alex Zarebski²⁴.

Samples and logistics:

Evelien M Adriaenssens⁵¹, Shazaad S Y Ahmad^{2,47}, Adela Alcolea-Medina^{59,77}, John Allan⁶⁰, Patawee Asamaphan²¹, Laura Atkinson⁴⁰, Paul Baker⁶³, Jonathan Ball⁵⁵, Edward Barton⁶⁴, Mathew A Beale¹, Charlotte Beaver¹, Andrew Beggs¹⁶, Andrew Bell⁵¹, Duncan J Berger¹, Louise Berry⁵⁶, Claire M Bewshea⁴⁹, Kelly Bicknell⁷⁰, Paul Bird⁵⁸, Chloe Bishop⁷, Tim Boswell⁵⁶, Cassie Breen⁴⁸, Sarah K Buddenborg¹, Shirelle Burton-Fanning⁶⁶, Vicki Chalker⁷, Joseph G Chappell⁵⁵, Themoula Charalampous^{78,94}, Claire Cormie³, Nick Cortes^{29,25}, Lindsay J Coupland⁵², Angela Cowell⁴⁸, Rose K Davidson⁵³, Joana Dias³, Maria Diaz⁵¹, Thomas Dibling¹, Matthew J Dorman¹, Nichola Duckworth⁵⁷, Scott Elliott⁷⁰, Sarah Essex⁶³, Karlie Fallon⁵⁸, Theresa Feltwell⁸, Vicki M Fleming⁵⁶, Sally Forrest³, Luke Foulser¹, Maria V Garcia-Casado¹, Artemis Gavriil⁴¹, Ryan P George⁴⁷, Laura Gifford³³, Harmeet K Gill³, Jane Greenaway⁶⁵, Luke Griffith⁵³, Ana Victoria Gutierrez⁵¹, Antony D Hale⁸⁵, Tanzina Haque⁹¹, Katherine L Harper⁸⁵, Ian Harrison⁷, Judith Heaney⁸⁹, Thomas Helmer⁵⁸, Ellen E Higginson³, Richard Hopes², Hannah C Howson-Wells⁵⁶, Adam D Hunter¹, Robert Impey⁷⁰, Dianne Irish-Tavares⁹¹, David A Jackson¹, Kathryn A Jackson⁴⁶, Amelia Joseph⁵⁶, Leanne Kane¹, Sally Kay¹, Leanne M Kermack³, Manjinder Khakh⁵⁶, Stephen P Kidd^{29,25,31}, Anastasia Kolyva⁵¹, Jack CD Lee⁴⁰, Laura Letchford¹, Nick Levene⁷⁹, Lisa J Levett⁸⁹, Michelle M Lister⁵⁶, Allyson Lloyd⁷⁰, Joshua Loh⁶⁰, Louissa R Macfarlane-Smith⁸⁵, Nicholas W Machin^{2,47}, Mailis Maes³, Samantha McGuigan¹, Liz McMinn¹, Lamia Mestek-Boukhibar⁴¹, Zoltan Molnar⁶, Lynn Monaghan⁷⁹, Catrin Moore²⁷, Plamena Naydenova³, Alexandra S Neaverson¹, Rachel Nelson¹, Marc O Niebel²¹, Elaine O'Toole⁴⁸, Debra Padgett⁶⁴, Gaurang Patel¹, Brendan Al Payne⁶⁶, Liam Prestwood¹, Veena Raviprakash⁶⁷, Nicola Reynolds⁸⁶, Alex Richter¹⁶, Esther Robinson⁹⁵, Hazel A Rogers¹, Aileen Rowan⁹⁶, Garren Scott⁶⁴, Divya Shah⁴⁰, Nicola Sheriff⁶⁷, Graciela Sluga, Emily Souster¹, Michael Spencer-Chapman¹, Sushmita Sridhar^{1,3}, Tracey Swingler⁵³, Julian Tang⁵⁸, Graham P Taylor⁹⁶, Theocharis Tsoleridis⁵⁵, Lance Turtle⁴⁶, Sarah Walsh⁵⁷, Michelle Wantoch⁸⁶, Joanne Watts⁴⁸, Sheila Waugh⁶⁶, Sam Weeks⁴¹, Rebecca Williams³¹, Iona Willingham⁵⁶, Emma L Wise^{25,29,31}, Victoria Wright⁵⁴, Sarah Wyllie⁷⁰, and Jamie Young³.

Software and analysis tools

Amy Gaskin³³, Will Rowe¹⁵, and Igor Siveroni⁹⁶.

Visualisation:

Robert Johnson⁹⁶.

1 Wellcome Sanger Institute, **2** Public Health England, **3** University of Cambridge, **4** Health Data Research UK, Cambridge, **5** Public Health Agency, Northern Ireland, **6** Queen's University Belfast **7** Public Health England Colindale, **8** Department of Medicine, University of Cambridge, **9** University of Oxford, **10** Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust; Cambridge, UK, **11** Division of Virology, Department of Pathology, University of Cambridge, **12** The Francis Crick Institute, **13** Cambridge Institute for Therapeutic Immunology and Infectious Disease, Department of Medicine, **14** Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK, **15** Institute of Microbiology and Infection, University of Birmingham, **16** University of Birmingham, **17** Queen Elizabeth Hospital, **18** Heartlands Hospital, **19** University of Edinburgh, **20** NHS Lothian, **21** MRC-University of Glasgow Centre for Virus Research, **22** Institute of Biodiversity, Animal Health & Comparative Medicine, University of Glasgow, **23** West of Scotland Specialist Virology Centre, **24** Dept Zoology, University of Oxford, **25** University of Surrey, **26** Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, **27** Big Data Institute, Nuffield Department of Medicine, University of Oxford, **28** Oxford University Hospitals NHS Foundation Trust, **29** Basingstoke Hospital, **30** Centre for Genomic Pathogen Surveillance, University of Oxford, **31** Hampshire Hospitals NHS Foundation Trust, **32** University of Southampton, **33** Public Health Wales NHS Trust, **34** Cardiff University, **35** Betsi Cadwaladr University Health Board, **36** Cardiff and Vale University Health Board, **37** Swansea University, **38** University of Sheffield, **39** Sheffield Teaching Hospitals, **40** Great Ormond Street NHS Foundation Trust, **41** University College London, **42** Oswaldo Cruz Institute, Rio de Janeiro **43** North West London Pathology, **44** Imperial College Healthcare NHS Trust, **45** NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, **46** University of Liverpool, **47** Manchester University NHS Foundation Trust, **48** Liverpool Clinical Laboratories, **49** University of Exeter, **50** Royal Devon and Exeter NHS Foundation Trust, **51** Quadram Institute Bioscience, University of East Anglia, **52** Norfolk and Norwich University Hospital, **53** University of East Anglia, **54** Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, **55** Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, **56** Clinical Microbiology Department, Queens Medical Centre, **57** PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, **58** Clinical Microbiology, University Hospitals of Leicester NHS Trust, **59** Viapath, **60** Hub for Biotechnology in the Built Environment, Northumbria University, **61** NU-OMICS Northumbria University, **62** Northumbria University, **63** South Tees Hospitals NHS Foundation Trust, **64** North Cumbria Integrated Care NHS Foundation Trust, **65** North Tees and Hartlepool NHS Foundation Trust, **66** Newcastle Hospitals NHS Foundation Trust, **67** County Durham and Darlington NHS Foundation Trust, **68** Centre for Enzyme Innovation, University of Portsmouth, **69** School of Biological Sciences, University of Portsmouth, **70** Portsmouth Hospitals NHS Trust, **71** University of Warwick, **72** University Hospitals Coventry and Warwickshire, **73** Warwick Medical School and Institute of Precision Diagnostics, Pathology, UHCW NHS Trust, **74** Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, **75** Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, **76** Department of Infectious Diseases, King's College London, **77** Guy's and St. Thomas' Hospitals NHS Foundation Trust, **78** Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, **79** Princess Alexandra Hospital Microbiology Dept. , **80** Cambridge University Hospitals NHS Foundation Trust, **81** East Kent Hospitals University NHS Foundation Trust, **82** University of Kent, **83** Gloucestershire Hospitals NHS Foundation Trust, **84** Department of Microbiology, Kettering General Hospital, **85** National Infection Service, PHE and Leeds Teaching Hospitals Trust, **86** Cambridge Stem Cell Institute, University of Cambridge, **87** Public Health Scotland, **88** Belfast Health & Social Care Trust, **89** Health Services Laboratories, **90** Barking, Havering and Redbridge University Hospitals NHS Trust, **91** Royal Free NHS Trust, **92** Maidstone and Tunbridge Wells NHS Trust, **93** University of Brighton, **94** Kings College London, **95** PHE Heartlands, **96** Imperial College London, **97** Department of Infection Biology, London School of Hygiene and Tropical Medicine.