This is a repository copy of *An integrated national scale SARS-CoV-2 genomic surveillance network*.

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/169457/

Version: Published Version

Article:
The COVID-19 Genomics UK (COG-UK) consortium (2020) An integrated national scale SARS-CoV-2 genomic surveillance network. The Lancet Microbe, 1 (3). e99-e100. ISSN 2666-5247

https://doi.org/10.1016/s2666-5247(20)30054-9

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 including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

An integrated national scale SARS-CoV-2 genomic surveillance network

Published Online
June 2, 2020
https://doi.org/10.1016/S2666-5247(20)30054-9

The Coronavirus Disease 2019 (COVID-19) Genomics UK Consortium (COG-UK) was launched in March, 2020, with £20 million support from UK Research and Innovation, the UK Department of Health and Social Care, and Wellcome Trust. The goal of this consortium is to sequence severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) for up to 230 000 patients, health-care workers, and other essential workers in the UK with COVID-19, which will help to enable the tracking of SARS-CoV-2 transmission, identify viral mutations, and integrate with health data to assess how the viral genome interacts with cofactors and consequences of COVID-19.

Results from this initiative are guiding decision makers, including the weekly reports to the UK Scientific Advisory Group for Emergencies (SAGE). This initiative is the first time that large-scale genomic epidemiology has been used to guide and inform the public health response to a pandemic in the UK, setting the stage for genomics to serve as a core tool for outbreak tracking in future pandemics.

COG-UK builds on the UK's strengths in pathogen genomics, population health sciences, and health informatics. It benefits from a large and well equipped network of specialist academic and research facilities working in close collaboration with the UK's public health agencies and the National Health Service (NHS). Partners in COG-UK at the time of writing are shown in the appendix (p 4). As COG-UK's example could help to inform other countries seeking to rapidly develop and scale-up national sequencing capacity and a joined-up health information system, we describe six key features of our experience.

COG-UK is co-ordinated from an integrated hub (based between the University of Cambridge [Cambridge, UK] and Wellcome Sanger Institute [Cambridge, UK]), with sample collection and sequencing taking place at multiple organisations across the country. This decentralised model enables rapid sequencing and prioritisation at the point of need, while supporting equitable national access. The core of COG-UK is formed around a network of regional sequencing centres in UK academic institutions and UK public health agencies. Sequencing in regional sequencing centres is close to real time, with a 24–48 h turnaround time, with data interpreted and used locally. In addition, the Wellcome Sanger Institute provides a high volume national sequencing capacity using high throughput, cost efficient viral sequencing for UK hospitals that do not have sequencing capabilities (ie, contributing partners), the national testing centres for key workers, and overspill from the regional sequencing centres.

For more on **COG-UK** see https://www.cogconsortium.uk

To help prioritise finite sequencing resources, COG-UK has developed a sampling strategy to concurrently enable broad population-level analyses, targeted analyses of specific populations, and freedom to tackle local priorities. This strategy aims to maximise insight while minimising demands on stretched local facilities.

To maximise speed and inclusiveness, COG-UK adopts a range of technical approaches (ie, amplicons, baits, and metagenomics) and technologies (ie, single molecule *vs* next generation sequencing approaches), enabling centres to build on existing pipelines and expertise. Several centres have adopted a tiling amplicon sequencing approach developed by the ARTIC network, which is run on multiple platforms to help achieve the very high sample throughput required. Tight linkage to diagnostic and public health laboratories minimises transport and analysis delays. The system needs to be continuous and rapid, with a target of 48 h from sample collection to analysis.

For more on the **ARTIC network** see https://artic.network

Sequence data are uploaded to the Cloud Infrastructure for Microbial Bioinformatics (MRC-CLIMB) server;[1] a centralised, replicated environment for data storage and analysis. MRC-CLIMB provides a ready-made starting point for the computational analysis, with the option to integrate with other resources within the UK or scale into the commercial cloud. A standardised lineage assignment to enable national and international comparison has been developed,[2] which are linked to data and released interactively via Microreact.[3] Sequence data are made open access through release into the European Nucleotide Archive through the European Bioinformatics Institute and the Global Initiative on Sharing All Influenza Data.

For more on **European Bioinformatics Institute** see https://www.ebi.ac.uk/ena

For more on the **Global Initiative on Sharing All Influenza Data** see https://www.gisaid.org

To enhance the value of sequence data, we are creating an integrated dataset connecting viral genome data with multidimensional patient data from clinical,

epidemiological, and other sources. By accessing existing NHS e-health records and related sources, the goal is to avoid adding to the burden on busy NHS staff.

Each week a sequence data cutoff is applied across the consortium to give a defined dataset for weekly analysis to report to SAGE, initially focusing on: (1) local transmission versus imported cases,[4] (2) rates of epidemic growth,[5,6] (3) reconstructing spatial movement, (4) chains of transmission,[7] (5) observed genetic changes, and (6) identification of genomic changes potentially affecting common diagnostic tests or direct (eg, chemotherapeutics) or indirect therapies.

See Online for appendix

Our website will provide a publication strategy, including open access publications, bespoke local and regional analyses, and data summaries suitable for the general public.

In only 4 weeks, over 7000 genomes have been sequenced by COG-UK, the largest number of any country to date. We are committed to open and global collaboration. Reciprocity is crucial to genomic epidemiological approaches; the use of our data will be maximised if other countries adopt similar approaches.

We declare no competing interests. COG-UK is supported by the funding from the UK Research and Innovation (UKRI), Medical Research Council, and the National Institute of Health Research and Wellcome, administered by UKRI.

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

*The COVID-19 Genomics UK (COG-UK) consortium
contact@cogconsortium.uk

*Full list of consortium names and affiliations are in the appendix.

1 Connor TR, Loman NJ, Thompson S, et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2016; **2:** e000086.

2 Rambaut A, Holmes EC, Hill V, et al. A dynamic nomenclature proposal for SARS-CoV-2 to assist genomic epidemiology. *bioRxiv* 2020; published online April 19. DOI: 0.1101/2020.04.17.046086.

3 Argimón S, Abudahab K, Goater RJE, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microb Genom* 2016; **2:** e000093.

4 Baillie GJ, Galiano M, Agapow P-M, et al. Evolutionary dynamics of local pandemic H1N1/2009 influenza virus lineages revealed by whole-genome analysis. *J Virol* 2012; **86:** 11–18.

5 Fraser C, Donnelly CA, Cauchemez S, et al. Pandemic potential of a strain of influenza a (h1n1): early findings. *Science* 2009; **324:** 1557–61.

6 Volz EM, Kosakovsky Pond SL, Ward MJ, Leigh Brown AJ, Frost SD. Phylodynamics of infectious disease epidemics. *Genetics* 2009; **183:** 1421–30.

7 Houlihan CF, Frampton D, Ferns RB, et al. Use of whole-genome sequencing in the investigation of a nosocomial influenza virus outbreak. *J Infect Dis* 2018; **218:** 1485–89.