

COVID-19 Genomics UK Consortium (COG-UK)

Weekly Report #2

1st April 2020

Executive Summary

- COG-UK has brought online an additional sequencing centre (Cambridge) and increased the number SARS-CoV-2 genomes sequenced and analysed to a total of 806 (up from 260 on the 23rd March). The UK has now reported the largest number of genomes of any individual country in the pandemic to date.
- Ongoing analyses support a model of a large number of independent SARS-CoV introductions into the UK from around the world, with at least 12 lineages in circulation discernable from the current dataset.
- The integration of epidemiological data with viral genome sequence data, while limited so far and in need of acceleration, has already begun to provide valuable insight into nosocomial (i.e. originating in a hospital) transmission events and informed frontline approaches in a manner that would not be possible with epidemiological data alone.
- While acknowledging the potential for conflict between diagnostic testing and genome sequencing capacity, adequate infrastructure to ensure sample flow, sequence centre staffing and supply of extraction kits and sequencing reagents are potential risks. Mitigating these issues will be important for scaling up COG-UK throughput to the levels needed to maximise the impact of SARS-CoV-2 genomic surveillance in the UK.

COG-UK update

In the past week, an additional sequencing centre at the University of Cambridge has been brought on line, bringing the total number of already active COG-UK sites to eight. A further six sequencing centres are expected to come online shortly (Figure 1).

By the data cut-off for this report, the total number of viral genomes now stands at 806 (Table 1). To place this in context, in the 21 days that the consortium has been active to date, COG-UK has generated more genomes than reported by any other country during the entire epidemic to date (Table 2).

From the genomes available, we have identified at least 12 viral lineages circulating within the UK. Continued under-sampling in the UK and elsewhere means that the number of independent introductions of SARS-CoV-2 likely remains substantially higher.

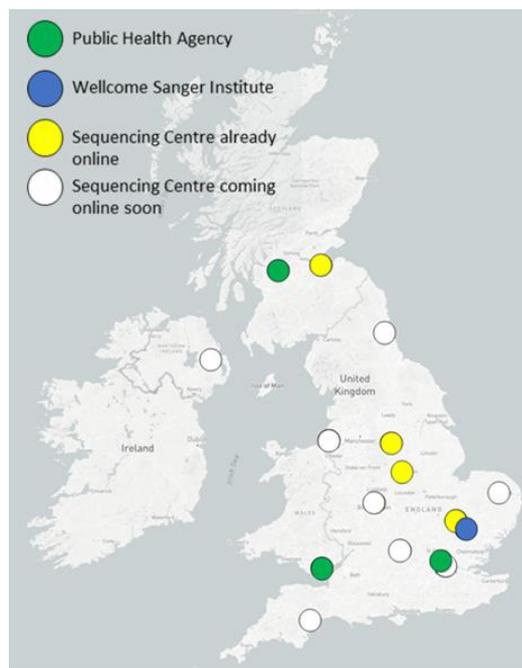


Figure 1: Location of current COG-UK sites.

Sequencing Centre	Number of SARS-CoV-2 genomes	
	22/03	31/03*
Cambridge	-	33
Edinburgh	32	86
Glasgow	27	102
Nottingham	1	28
Public Health England (Colindale)	44	147
Public Health Wales (Cardiff)	48	218
Wellcome Sanger Institute	5	143
Sheffield	13	49
TOTAL	260	806

Table 1: Number of SARS-CoV-2 genomes sequenced and analysed by the COG-UK centres. *includes 33 duplicate genomes generated while bringing Cambridge sequencing centre online

Country	Number of SARS-CoV-2 genomes		
	11/03	22/03	31/03
UK	24	260	806
USA	70	182	744
Iceland	0	0	343
China	206	280	296
Netherlands	28	190	190
Canada	3	17	129
France	10	61	119
Japan	10	12	84
Australia	18	30	71
Spain	0	3	42
Germany	3	15	27
Italy	5	5	23

Table 2: Number of SARS-CoV-2 genomes sequences reported in GISAID. Twelve highest countries shown.

Genomics data generated in Wales as part of COG-UK has been used to provide evidence of nosocomial transmission within a hospital in Wales. This result has been fed into the incident management team within the health board concerned (see Analysis section below for more detail).

The COG-UK website (<http://cogconsortium.uk>) will be made publicly available from 1st April.

Recommendations

- Genomic data alone in the absence of sufficiently detailed metadata will be of limited use when drawing conclusions and making inferences about the spread and evolution of SARS-CoV-2 in the UK. It therefore remains imperative that we work towards gaining access to patient electronic health records and aggregated mobility data to enable the incorporation of near real-time epidemiological information into our analyses, and to understand and inform measures to control viral introduction and transmission.
- In addition to cases for which known epidemiological links are corroborated by genomic data, we can now begin to search for clusters in the genomic data that could be marked for further epidemiological investigation as linked cases, perhaps indicative of nosocomial transmission or incidence of super-spreading.
- Adequate infrastructure to ensure sample flow and suitable capacity in sequencing centres in the network is a potential limiting factor in scaling up COG-UK throughput to the levels needed to maximise the impact of SARS-CoV-2 genomic surveillance. Ensuring sufficient support is provided for each centre in the COG-UK network will be crucial.
- Furthermore, a lack of UK manufacturers for the RNA extraction kits and sequencing reagents required places at risk the ability for COG-UK sequencing centres to keep pace with need. Ensuring continued supply of kits and reagents either by developing alternatives locally or by working closely with manufacturers elsewhere to secure appropriate supply lines is a priority.

Analysis

Evidence of nosocomial transmission in a Welsh hospital

Early analysis of sequenced Welsh samples has identified four large SARS-CoV-2 clusters. The largest of these, comprising 72 samples as of 30/03/2020, was identified as being linked to a hospital within a Welsh health board. This cluster forms a discrete group within a larger cluster comprising viruses from Brazil, Germany and the Netherlands. On the 26th of March a request from Public Health Wales (PHW) Health Protection was made to perform a preliminary analysis of the cluster to establish whether a number of cases associated with this cluster were consistent with nosocomial transmission.

The earliest sample within the cluster had been identified as a putative index case by the PHW Healthcare Associated Infection team. This patient was found to be infected by SARS-CoV-2 when routine testing of ICU patients was introduced.

Using a list of 27 contact-traced healthcare workers (HCWs), patients and familial/household contacts linked to the putative index case, we were able to:

- Identify that all 27 cases fell within this cluster
- Identify that all cases carried an essentially identical virus to the index case
- Establish that the index case sat at the base of the group

In 7 cases there was no known contact between the HCW and the patient, although shift overlaps with HCWs who did have patient contact were found for 3 of these cases. This finding may imply nosocomial transmission beyond the immediate contact with the index case patient, which would be consistent with the patient being initially undiagnosed with COVID-19 and is suggestive of potential transmission via surface contamination.

Collectively, genomics combined with epidemiological data has identified a cluster of cases that is consistent with nosocomial transmission. Within the context of a larger background of cases within a health board, the use of genomics has enabled the rapid identification (the initial analysis was returned to the PHW team within a couple of hours) of a clear group of related cases, which can then be examined/explained using epidemiological data. This information is already being used for incident management by the health board concerned. A more extensive summary of this work is being produced and will hopefully be provided to SAGE in the near future.

What can we learn from genomics about nosocomial spread?

A detailed analysis of SARS-CoV-2 in Iceland published this week demonstrated the added value of analysing contract tracing and genomic data together, and also reported cases imported to Iceland from the UK (<https://www.medrxiv.org/content/10.1101/2020.03.26.20044446v2>). As with the above example described for the hospital cluster in Wales, integration of epidemiological data with viral genome sequence data can provide valuable insight in several ways.

1) Supporting contact tracing:

Although the SARS-CoV-2 genomes are not very diverse, observing the same (or nearly identical) genome in the putative index case and in his/her contacts can corroborate the epidemiological contact tracing. On their own (i.e. without the contact data), viral genomes cannot prove or demonstrate a connection. However, the genomes can reject a putative epidemiological link. That is, if a contact has a genome that is sufficiently different from the index case genome (or from a cluster), then we can conclude with confidence, from the genomes alone, that there is no direct epidemiological connection. Exactly what “sufficiently different” is needs to be quantified and defined. Since at this stage in the epidemic reversion of mutation is very rare, we can assume that if a contact does not share one of the mutations that defines the index case/cluster, then they are not epidemiologically connected.

2) Population level estimates of nosocomial transmission:

Population level estimations have not been undertaken yet, but will be based on phylodynamic models. They will provide estimates of the relative contribution of nosocomial spread at a national scale, but will not give information about individual hospitals. The COG-UK analysis group is already in discussion about some of the ways in which these kinds of analysis could be undertaken. Our initial conclusion is that to make population estimates, we will need sufficient sampling from both community and hospitals, and appropriate metadata (i.e. are the genomes from HCW, hospitalised patients, or from people in the community, as well as the location at local authority level, rather than county level).