Utility of whole genome sequencing for SARS-CoV-2

Background

Whole Genome Sequencing (WGS) allows viruses to be genetically fingerprinted, providing greater resolution to distinguish the viruses causing disease beyond current diagnostic testing methods. As SARS-CoV-2 spreads throughout Scotland, the capacity for real-time viral WGS offers the opportunity to track SARS-CoV-2 disease throughout the country. In a relatively small country like Scotland, this offers the ability to better understand the number of introductions and impact of distancing measures on virus spread in our communities. It also provides a tool to help detect and contain nosocomial transmissions in hospitals and nursing homes, better protect our communities and health care workers (HCW), and guide hospital bed-flow management.

Whole genome sequencing of SARS-CoV-2 in Scotland

Genome sequencing of the COVID-19 virus is being undertaken by a Glasgow and Edinburgh partnership working as part of the COVID-19 Genomics UK (COG-UK) Consortium supported by the UK Government (www.cogconsortium.uk). WGS data generated by the consortium is being shared with Public Health Scotland (PHS) and NHS partners for public health and patient benefit, and genomics has been integrated as part of the national diagnostic strategy.

The COG-UK partner laboratories in Glasgow and Edinburgh laboratories have contributed over 1300 SARS-CoV-2 sequences to the UK consortium using Oxford Nanopore GridION real-time sequencing platforms that rapidly sequence samples. The turnaround time from sample reception to whole genome sequence data is currently 2 days, and means that it is possible to produce actionable data and results within a clinically relevant time frame. The cost to sequence a sample is £55.

What can this genome data be used for?

Scotland has world lead research expertise in viral genomics and epidemiology. Scottish COG-UK are harnessing the strength this fundamental science to translate it into the COVID-19 response, generating results which are useful for the people of Scotland, and the leadership of Scotland. Real-time SARS-CoV2 sequencing is feeding into the national surveillance data and can contribute towards:

- Detection of within country movement of SARS-CoV2, and characterisation of localised clusters
- Distinguishing between local/community transmissions and imported cases
- Detecting chains of transmission, in particular in hospital and care home settings
- Tracking of the impact of intervention measures such as the Test, Trace, Isolate, Support (TTIS) strategy
- Estimation of reproduction number, and number of unreported cases from viral phylodynamic and epidemiological models
- Monitoring the viral population for mutations in the diagnostic test targets that may impact on test sensitives
- When future therapeutic interventions are available, mutations in viral proteins that reduce the efficacy can be detected, and real-time sequencing can inform clinical management

Targeting sampling and sequencing to support the COVID-19 response

In collaboration with the Scottish Microbiology and Virology Network (SMVN) and the NHS diagnostic laboratories, samples from throughout Scotland are available for sequencing. The Scottish NHS Boards have been divided between Edinburgh and Glasgow sequencing centres to ensure complete national coverage.

Sequencing is being carried out in two tiers, with related but distinct aims:

Tier 1 – **population-level** epidemiological sampling: relative proportion of cases from all health boards based on population. Aim - Build up a Scotland-centric baseline of SARS-CoV-2 spread, virus genomic epidemiology and inform the national policy on COVID-19 control measures.

Tier 2 – **healthcare level** local prioritisation: Intensive Care Unit (ICU) patients, patient and HCWs with suspected nosocomial infections (hospital and nursing homes), and community infections. Aim - Use WGS as a diagnostic tool to identify local transmission clusters within hospitals and nursing homes and inform IPC measures to better protect both patients and health care workers.

Using WGS to unravel the origins of the epidemic spread of COVID-19 in Scotland

The spread of the virus around the global has been swift, with the first recorded case of COVID-19 detected in Scotland on the 1st of March. Since then, there has be a rapid epidemic rise in cases across Scotland. Using genome sequencing data it has been possible to reconstruct the origins of the outbreak in Scotland, and demonstrates there have been multiple introductions of the virus into Scotland (Figure 1). Several of these introduced have led to further transmission clusters associated with various settings including residential facilities (Figure 2). Monitoring the 'success' of these lineages and as data accumulates the size of the infected population will be a key contributor to quantifying the impact of control measures. In addition, as social distancing measures are relaxed, further sequencing of cases has the potential to detect cryptic transmission routes and novel introductions of the virus into Scotland.

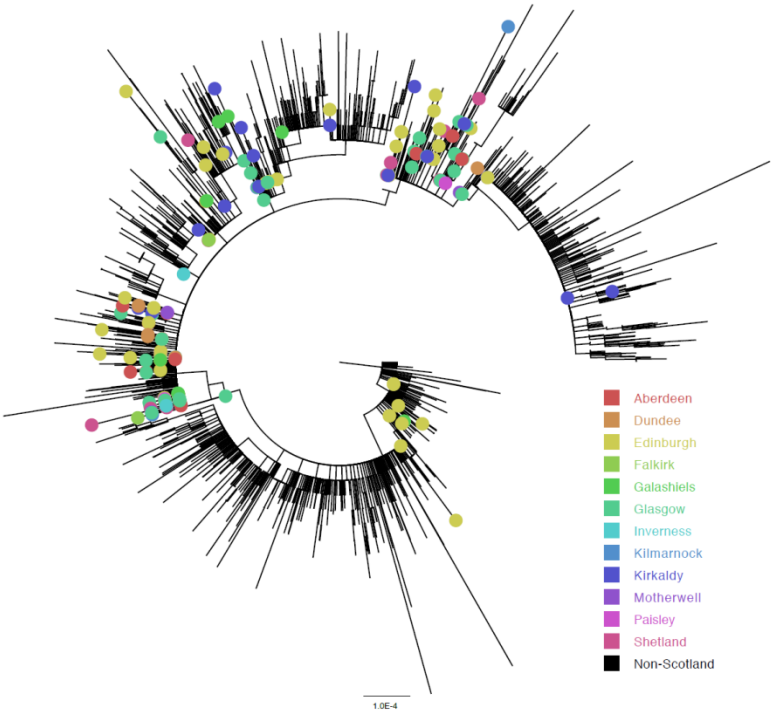

Figure 1: Phylogeny of global SARS-CoV-2 sequences with Scottish isolates highlighted by coloured dots corresponding to postcode area demonstrating multiple introductions across Scotland.

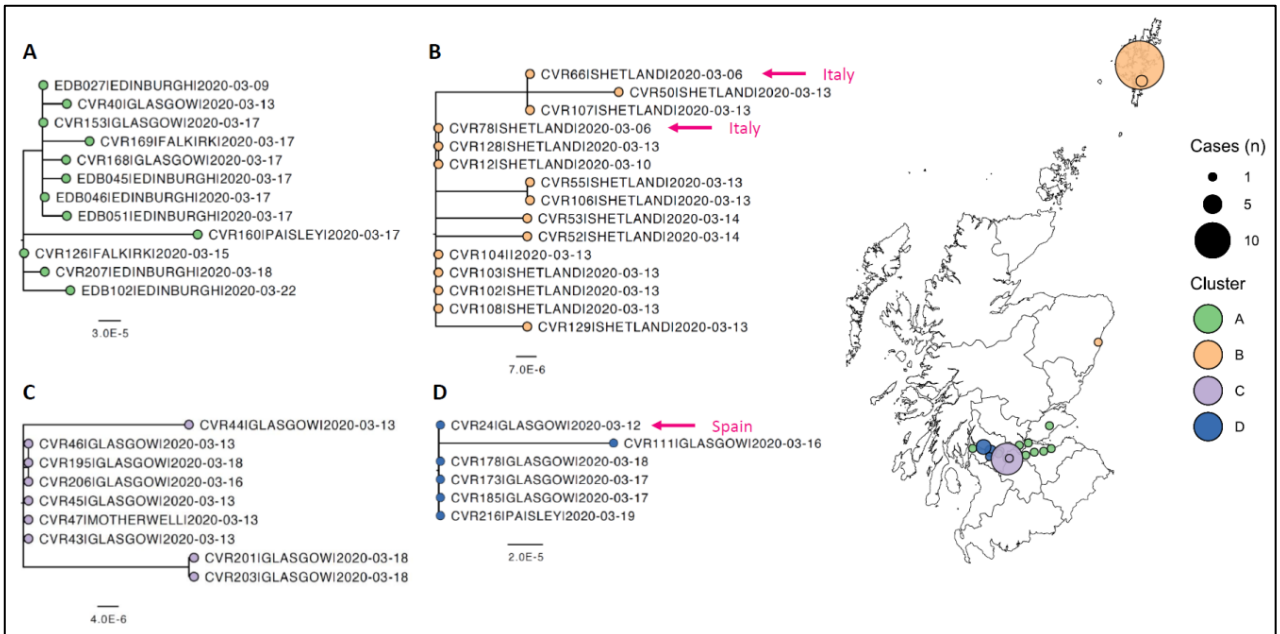

Figure 2: Spatial association of phylogenetic clusters associated with introduction events; (A) A spatially distributed outbreak without an obvious link to travel. (B) A local cluster demonstrating travel associated introduction into a geographically restricted area by two linked cases. (C) A focal outbreak in a residential facility not associated with any known travel. (D) A cluster associated with travel to Spain.

Using WGS to investigate nosocomial infections

Within healthcare settings genome sequences can be used to 'rule out' transmissions, where viruses are clearly not genetically linked, and also potentially 'rule in' transmissions where sequences are indistinguishable or very close related. The ability to distinguish between the two will allow resources to be targeted appropriately, and infection prevention and control (IPC) measures to be employed effectively. The utility of genomic epidemiology to inform infection control at high resolution is highlighted by the case of a healthcare worker (HCW) who tested positive for CoVID-19 whilst working on a hospital ward managing patients with confirmed COVID-19. The HCW-derived sequence was clearly distinct to those from the two COVID-19 patients that had been managed on the ward (Figure 3), suggesting nosocomial acquisition highly unlikely. Such information can be used to inform targeted infection control interventions and monitor the real-world effectiveness of personal protective equipment (PPE). WGS to investigate a suspected nosocomial outbreak of COVID19. Infected HCWs can be contributors to virus spread, for example, in hospital wards. A study has been instigated to sample all patients on a ward and compare their sequences to those of HCWs known to have been found to be SARS-CoV-2 positive at a later date. This type of analysis can help monitor for nosocomial outbreaks.

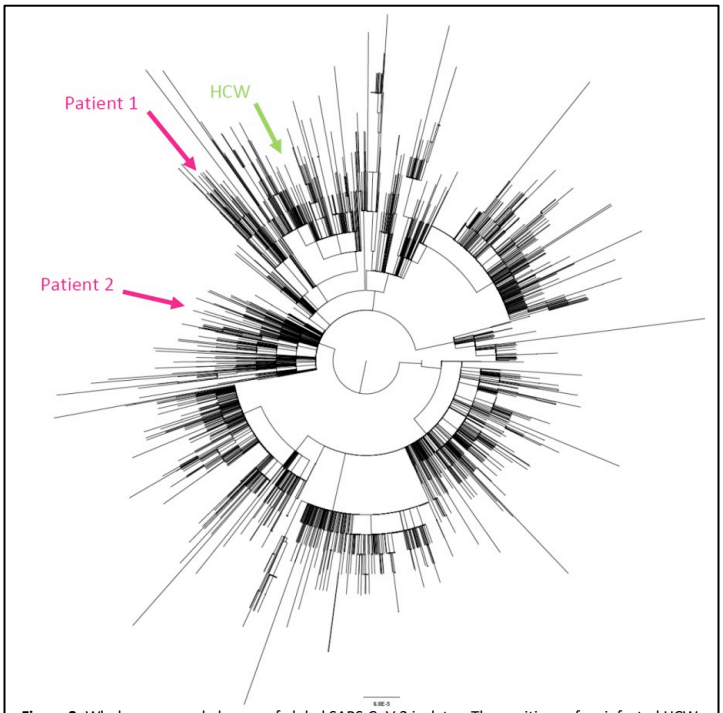

Figure 3: Whole genome phylogeny of global SARS-CoV-2 isolates. The positions of an infected HCW and two inpatients with whom they had contact is indicated by arrows. In this case there was sufficient genetic distance between the isolates to exclude cross-transmission.

Building up a baseline picture of the SARS-CoV-2 population and future surveillance

The continued sequencing of SARS-CoV-2 from the first reported case, through the lockdown to the current day, is building up a baseline picture of the virus population, which we will be able to use to monitor and measure the effectiveness of future interventions and policy strategies.

Genome sequencing is providing a high resolution picture of the population of SARS-CoV-2 circulating Scotland. COG-UK team members have produced a typing scheme that can be used to divide up the population into different lineages (https://www.biorxiv.org/content/10.1101/2020.04.17.046086v1), and a bioinformatic tool that can do this rapidly from the WGS data (Pangolin: https://github.com/hCoV-2019/pangolin). The ability to break down the population structure based on genetic variants is an invaluable tool for the public health response. It means that lineages can easily be identified and followed over time to see how they change and spread.

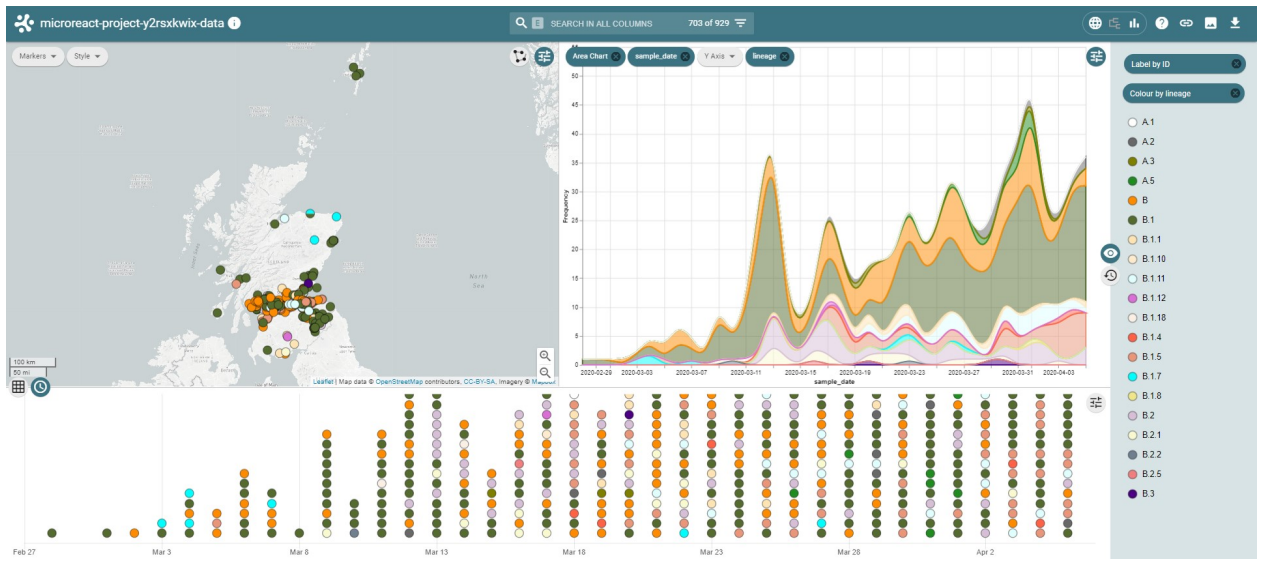

Figure 4: Microreact tool displaying a national picture of the linked SARS-CoV-2 genome and epidemiological data. The dots represent SARS-CoV-2 samples and are colour coded according to the SARS-CoV-2 lineages that are circulating in the Scotland (see left hand panel for the different types). The map view of the home residences of the patients (top left) displays the geospatial distribution of cases; the overlapping line plot (top right) illustrating cumulative frequency of the different SARS-CoV-2 lineages over time across Scotland, and the time line view (bottom) displays the temporal distribution of samples.

In partnership with COG-UK members, genome sequences generated by COG-UK from have been loaded into the Microreact web tool (https://microreact.org/) within PHS (Figure 4), and are helping to inform the COVID-19 response going forward; data from WGS is feeding into the Enhanced Surveillance and Modelling and Research Cells within PHS. In total 20 distinct global lineages that have been identified in Scotland and some of these lineages have become more dominant over time during the epidemic spread in across the country (Figure 4; see the overlapping plot, top left).

Comparing the Scottish WGS data to a wider UK and global collection, it has been possible to identify sub-lineages within the main lineages that are the results of multiple separate instructions into Scotland. Using the temporal, special and epidemiological information displayed in Microreact it is possible to identify and trace the spread of these introduction and subsequent localized transmission. Figure 5 illustrates one such sub-lineage identified in the West of Scotland. The patient who provided the first sample had a recent foreign travel history. The rest of the cases linked to this sub-lineage are clustered around the Greater Glasgow area, with a concentration of cases in West Dunbartonshire. Using this data in combination with TTIS strategy, it will be possible to follow and monitor the success of interventions to control and eradicate viral lineages across Scotland.

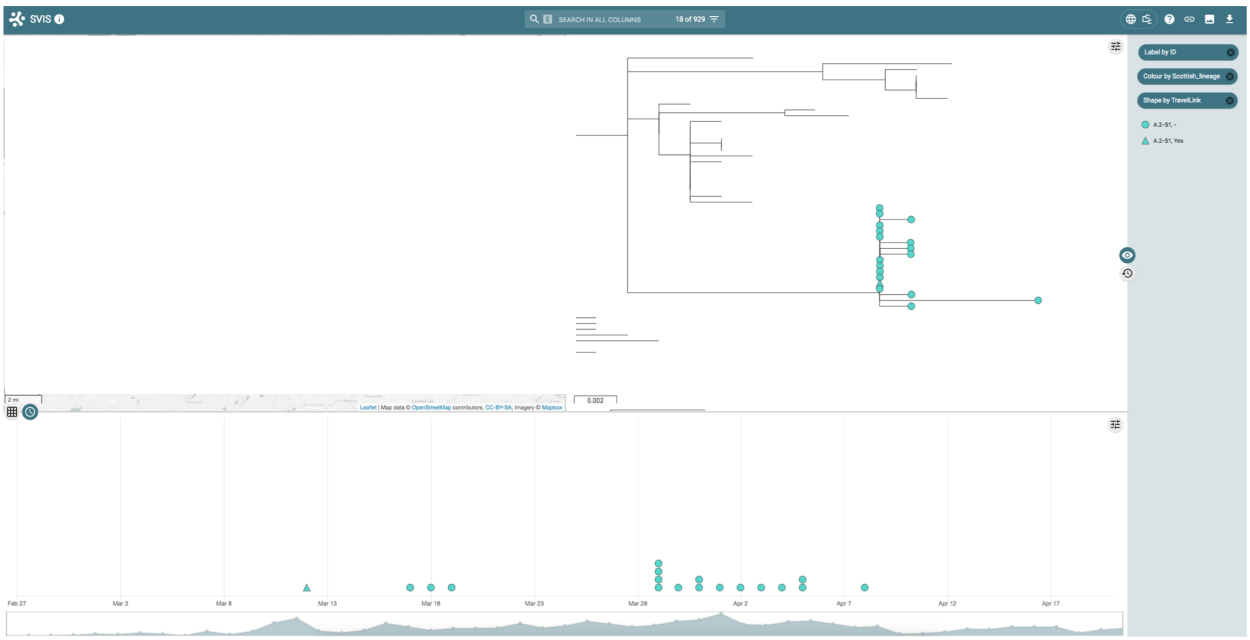

Figure 5: Microreact tool displaying the genetic relatedness and geographic distribution of a single lineage introduced to Scotland. The map view of the home residences of the patients (top left) displays the geospatial distribution of cases. The phylogenetic tree view illustrates the genetic relationships for the viruses (top right), where all the samples belonging to the same lineage are displayed as coloured dots or triangles (turquoise; where a triangle indicated a known foreign travel link, and circle indicated uncertain or no know travel link); and the time line view (bottom) displays the temporal distribution of samples.

Data Linkage to wider patient datasets

Within PHS WGS data is linked to the Electronic Communication of Surveillance in Scotland (ECOSS) database that contains patient information including: Community Health Index (CHI), date of birth, postcode of residency, sex, age, sample type, date sample taken, source in hospital, travel history and other clinically information. In addition, the genome data will be linked via CHI to other sets of patient data available in PHS that will allow the virus genomes to be compared to case severity and outcome, and co-morbidities and medication. These data linkages will enable associations between the virus genotype and disease to be investigated. This will allow horizon scanning for mutations that have impact on COVID-19 disease and transmission, and in turn can feed into clinical management and policy decisions.

Increasing the sequencing capacity to support the COVID-19 response

The activity of COG-UK on Scotland is dependent on funding from the UK government. COG-UK partners in Scotland are funded to sequence 300 SARS-CoV-2 samples per week, and this funding is secured until the end June, when it will be reassessed by the consortium. To secure longer term WGS capacity within Scotland, we are seeking funding from the Scottish Government to support the SARS-CoV-2 sequencing until to the end of the year. The funding will provide sequencing for 300 SARS-CoV-2 genomes per week. In addition, the funding request includes IT infrastructure within PHS, to create storage and analysis capacity that can be used to integrate SARS-CoV-2 genome data into NHS and PHS workflows and decision making. As our response to the virus becomes more nuanced with routine testing, return to work policies, and the availability of effective medicines, it will continue to be important to track the circulating SARS-CoV-2 population. The investment in WGS of SARS-CoV-2 in Scotland will provide this capacity, and help target our future interventions and response to COVID-19.

The Scottish COG-UK SARS-CoV-2 lead partners

MRC Centre for Virus Research, University of Glasgow: Emma Thomson & David Robertson; *University of Edinburgh:* Andrew Rambaut *NHS Lothian:* Kate Templeton; *NHS GG&C:* Rory Gunson; *Public Health Scotland:* Matthew Holden