Proposal for a National SARS-CoV-2 Surveillance & Whole Genome Sequencing Programme

Organisation: SARS-CoV-2 Surveillance & Whole Genome Sequencing Working Group

Date: 27 January 2021

Action required:
☐ For noting
☐ For discussion
X For decision

Approved for future publication: NO
Proposal for a National SARS-CoV-2 Surveillance & Whole Genome Sequencing Programme

27 January 2021

Table of contents

Development of the proposal .................................................................................................................. 3
Recommendations..................................................................................................................................... 3
Aims of the programme ............................................................................................................................ 7
Sampling strategy ....................................................................................................................................... 8
Capacity requirements .............................................................................................................................. 11
Data flow ................................................................................................................................................. 11
Governance .............................................................................................................................................. 12
Appendix 1: Membership of the Working Group ..................................................................................... 13
Appendix 2: Terms of reference of the Working Group ............................................................................ 14
Appendix 3: Background information ..................................................................................................... 15
References ............................................................................................................................................... 21
Development of the proposal

This proposal was developed by the members of the Working Group (Appendix 1) in response to a request from the Chief Medical Officer, Chair of NPHET on 6 January 2021. The proposal, which was developed in accordance with the agreed terms of reference (Appendix 2), was informed by input from the members of the Working Group, who represent a range of key stakeholders. The Working Group met four times, between 12 and 26 January 2021. To inform and support the development of this proposal, a rapid review to identify international guidance documents was conducted by HIQA on 14 Jan 2021. The background information that provides the context for the National SARS-CoV-2 Surveillance & Whole Genome Sequencing Programme is detailed in Appendix 3.

It is acknowledged that in parallel with the development of this proposal, considerable work has been undertaken by a number of clinical and academic institutions, including the National Virus Reference Laboratory (NVRL) and the Irish Coronavirus Sequencing Consortium (ICSC), to build capacity for whole genome sequencing (WGS) for SARS-CoV-2 in Ireland. Work has also been undertaken by the Health Protection Surveillance Centre (HPSC) to identify how national WGS capacity can be optimally used to enhance the national and targeted public health responses.

Recommendations

The Working Group has generated the following recommendations in relation to the establishment of a National SARS-CoV-2 Surveillance & Whole Genome Sequencing Programme (the National Programme):

Structure and aims

1. The National Programme should be led by the HPSC, with the NVRL as the lead diagnostic partner, with logistical support from, and under the oversight of HSE laboratory operations. This arrangement provides a legal framework for the collection of patient samples and data for surveillance purposes and leverages working arrangements and data flow pathways already in use for other virus surveillance programmes.

2. The overarching aim of the National Programme is to use WGS to inform the urgent public health response to the COVID-19 pandemic. The Programme should be structured according to two parallel streams of work (i.e. Stream 1 – Proactive Sequencing and Stream 2 – Reactive Sequencing). The prioritisation of areas within these Streams should be subject to ongoing monitoring based on the needs of the public health system. In the short term, prioritised activities within Stream 1 will include national surveillance of all variants (in community & hospital settings) and early detection of variants of concern from other countries, while prioritised activities within Stream 2 will include investigations of complex outbreaks, vaccine escape/reinfections, unexpected changes in transmissibility/virulence and antiviral resistance.

3. The target for the National Programme should be to routinely sequence approximately 10% of cases per week. The recommended short-term goal (6-12 weeks) should be to build to a target capacity of 1,200-1,500 sequences per week, with the medium-term (12-24 weeks) potential for a surge capacity of 5,000 sequences per week.
4. Consistent with guidance developed by the World Health Organisation (WHO), the National Programme should be adequately resourced in terms of staffing and equipment from the outset, such that sequencing does not disrupt routine diagnostic or public health activities. Both the Proactive and Reactive Streams (Streams 1 and 2) should be resourced adequately to allow the currently prioritised areas in each stream to be undertaken simultaneously, while having systems in place for surge capacity.

5. Non-genomic surveillance that facilitates early identification and potential triage of samples with a higher probability of being associated with a variant of concern (e.g., screening for S gene drop-out using the ThermoFisher TaqPath assay) should continue to be used alongside WGS.

**Governance and strategy**

6. The HPSC should establish governance structures, comprising involvement of relevant expertise (including surveillance, public health, laboratory sequencing, bioinformatics and infectious diseases as appropriate), in line with their usual processes for operating national surveillance programmes.

7. The HPSC should lead on development of a national surveillance strategy and implementation plan for the medium and long term (beyond 3 months).

8. Monitoring and evaluation of the implementation of the National Programme should be undertaken by the Office of the National Clinical Director for Health Protection, within the HSE, to ensure that the public health objectives are attained and that the use of WGS is prioritised to add maximum value to the public health response.

9. An external evaluation should be undertaken after 9 months to assess the public health benefits delivered and assess value for money.

**Short term capacity**

10. In the short term (within 3 months), additional surveillance and public health capacity within the HPSC will be required to ensure sufficient expertise is available to guide development of the National Programme, to interpret the data from the Programme and to address the public health objectives outlined. Funding for this capacity building will be required.

11. In light of the immediate need to expand national genomic surveillance for the health service as outlined in this proposal, the HSE should work with the HPSC and the NVRL to build capacity and ramp up sequencing activity in a robust and sustainable manner, as a matter of urgency. The NVRL and HSE should actively engage with all available resources, including publicly funded capacity (Irish Coronavirus Sequencing Consortium (ICSC) and the European Centre for Disease Prevention and Control (ECDC)) as well as national and international commercial capacity as needed.

12. The NVRL, as the lead diagnostic partner, should outsource sample preparation, sequencing and analysis activities, as needed, to meet the capacity required over and above its own activity. The NVRL must be satisfied that any other laboratory conducting WGS on behalf of the National Programme has adequate quality management systems in place, and that data protection and other legal and ethical considerations including in relation to patient confidentiality and access to data are adhered to. This engagement and sub-contracting will be conducted in accordance with usual governance structures and in compliance with HSE procurement protocols, with particular consideration given to quality, efficiency, capacity, expertise, turn-
around time and cost in selecting sub-contracted laboratories. (The NVRL, with the support of the HSE, has planned to reach this primary target of 1,200 to 1,500 sequences per week, over a 2 to 3 month time frame, by leveraging existing publicly funded capacity as well as contracting with commercial providers.)

13. All contractual arrangements will need to be approved by the HSE to ensure that running costs of increased sequencing activity can be met.

Long term capacity

14. Funding for additional resources for the medium- to longer-term (beyond 3 months) will be outlined in the implementation plan for the strategy that the HPSC will develop.

15. The HSE should continue its work with the NVRL (as previously recommended by NPHET) to build sustainable and flexible national sequencing capacity for the health service leveraging available publicly funded capacity (ICSC and ECDC) as well as national and international commercial capacity as needed, in order to meet the needs of existing and future national surveillance programmes.

Inputs

16. In accordance with its existing structures, the HPSC should develop and communicate the minimum reporting requirements regarding WGS specific for each setting (e.g., community vs. hospital) and context (e.g., national surveillance vs. vaccine-escape investigation). CIDR should be updated to reflect these reporting requirements. All reporting should be consistent with the usual requirements to safeguard patient confidentiality and minimise patient harm.

17. The HPSC in conjunction with the NVRL and regional Departments of Public Health should outline the request process for Stream 2 Reactive Sequencing, and develop prioritisation criteria for sequencing, especially regarding outbreaks and clusters. The HPSC should develop guidance on sampling of outbreaks for the purpose of sequencing to aid public health teams to gain the most benefit from phylodynamic analysis.

18. The sampling frame for national surveillance should be subject to ongoing monitoring and adjusted, as necessary, by the HPSC with input from Departments of Public Health, the NVRL and the Irish Epidemiological Modelling Advisory Group (IEMAG), as appropriate.

19. The HSE and HPSC should communicate to sequencing laboratories: the legal requirements; data handling requirements; the processes involved and data elements for notification of variants of concern. Following this, all sequencing laboratories should retrospectively review their genomic data and submit all relevant data to the HPSC as soon as possible. These laboratories must continue to notify the HPSC, of all variants of concern as soon as they are identified.

20. Local arrangements between hospitals and sequencing laboratories may continue for local WGS investigations, however laboratories must notify the HPSC of all detected variants of concern, in line with their legal obligations.

Outputs

21. The Clinical Director of the NVRL should be responsible for reporting sequence results to the HPSC, for all work it undertakes or subcontracts on behalf of the
National Programme to the HPSC, in accordance with the minimum dataset as specified by the HPSC.

22. The HPSC should provide the IEMAG with the latest sequencing data on a weekly or monthly basis to inform statistical modelling.

23. The HPSC, the NVRL and regional Departments of Public Health should agree a process to communicate sequencing results to relevant stakeholders (including patients).

24. The HPSC should work with the NVRL to ensure that sequencing data are shared with the national and global community to improve the national, regional and global public health response to SARS-CoV-2.
Aims of the programme

The overarching aim of the National SARS-CoV-2 Surveillance & Whole Genome Sequencing Programme is to use WGS to inform and enhance the urgent public health response to the COVID-19 pandemic.

This can be achieved through 10 public health objectives:

1. National surveillance of all variants (community & hospital)
2. Early detection of variants of concern which have originated from other countries
3. Monitoring emerging lineages in animals
4. Complex outbreak investigation
5. Vaccine escape/ reinfection investigation
6. Unexpected change in transmissibility/ virulence investigation
7. Antiviral resistance investigation
8. Human-animal transmission investigation
9. Testing immunocompromised patients undergoing antibody therapy
10. Unexpected change in diagnostic performance investigation.

Given the broad range of public health applications and the differing sampling approaches required for each, there is a need to prioritise objectives according to two parallel streams of work (Stream 1 – Proactive Sequencing, Stream 2 – Reactive Sequencing) (Fig. 1). It is important that both streams of work are resourced adequately to allow currently prioritised areas in each stream to be undertaken simultaneously, while having systems in place for surge capacity. The prioritisation of areas should be subject to regular review and ongoing monitoring based on the needs of the public health system.

Although the primary role of the National Programme is to optimise public health surveillance for SARS-CoV-2, the Programme recognises the need to share data with the global community to maximise outcomes for public health benefit. The potential role of the National Programme for research purposes should be considered once the Programme has been established.
Figure 1. Two parallel streams of the National Programme

### Sampling strategy

The sampling strategy for Streams 1 and 2 will differ due to the different public health objectives for each. Analysis of sequences should also be kept separate to avoid the non-random sampling of Stream 2 biasing the interpretation of the national surveillance in Stream 1.\(^{(3)}\)

**Stream 1:**

For national surveillance purposes, virus samples from different geographic locations and time points, as well as from patients with varied demographics and across the disease severity spectrum should be selected for sequencing.\(^{(1)}\)

**National surveillance in the community** is required to assess the prevalence of variant viruses in the population and also for the early detection of variant viruses.
The number of sequences required is proportional to the incidence of COVID-19, and so the demand for sequencing will fluctuate depending on the stage of the pandemic in Ireland. In line with European Commission recommendations, the National Programme will aim to sequence 5-10% of all SARS-CoV-2 RT-PCR positive test results.\(^{(5)}\) As of 24 January 2021, based on a 7-day total of 14,877 cases nationally,\(^{(6)}\) this equates to approximately 744-1,488 sequences per week. The number of weekly sequences required to achieve detection and quantification of variants among other circulating variants will be informed by the latest ECDC guidance.\(^{(1)}\) At a minimum, the National Programme will aim to be able to quantify with low precision a variant with an expected prevalence of 2.5% among all circulating variants. Based on the current incidence of COVID-19, a minimum of approximately 544-575 sequences will be needed nationally per week to achieve this goal.\(^{(1)}\) This volume would also allow detection of a variant where prevalence is as low as 0.5%.

Community samples processed by NVRL will be selected randomly for sequencing according to the sampling frame. Only samples with sufficient viral load and volume should be selected for sequencing.

The sampling frame will be determined by Public Health need and initially will be based on biologically significant age bands and community healthcare organisations (CHO). The sampling frame will be monitored and adjusted accordingly by the HPSC, with input from IEMAG, as required.

**Hospital surveillance** will be undertaken for the detection of the emergence of more virulent strains in patients with a severe disease course. Leveraging off the current approach to hospital surveillance for influenza, the HPSC will require hospital laboratories to select 5 to 10% of samples obtained on admission (or on diagnosis) from patients who are immunocompromised and or who have severe disease, for example patients who progress to non-invasive ventilation (NIV), ICU, death. Samples with sufficient viral load and volume should be selected and provided to the NVRL for sequencing.

**Early detection of variants from other countries** which may displace current in-country variants. As new variants of concern emerge internationally, it is important that there is sufficient capacity (and surge capacity, if required) within the Programme to be able to sequence samples from people arriving from the affected countries at short notice. Should it not be possible to sequence all samples, consideration should be given to sequencing a random sample. Where possible, consideration should be given to utilising non-genomic surveillance of all samples, for example, with regards to B.1.1.7 lineage (i.e. UK variant) screening for S gene drop-out using the ThermoFisher TaqPath assay or detecting the N501Y amino acid change using allele specific PCR for the B.1.351 lineage (i.e. South African variant), and reflexing samples for confirmatory WGS. The current recommendation to test all individuals entering Ireland, from a country of concern, should facilitate this.

While not considered a priority in the initial phase of the Programme, in the medium to longer term (> 3 months), consideration should be given to proactively monitoring emerging SARS-CoV-2 lineages within wild/domestic/farmed animal populations, as these lineages may have an impact on human health, such as those detected in minks.\(^{(7)}\) It may also be important to sequence a sample of cases in individuals that work closely with animals, especially if there are any unusual epidemiological features.
**Stream 2:**

The WHO recommends prioritising sequencing for the following situations:

- during cluster investigations when sequencing can support understanding of transmission events and/or evaluate the efficacy of infection control procedures
- in individuals vaccinated against SARS-CoV-2, but who later become infected with SARS-CoV-2 despite exhibiting an appropriate immune response to the vaccine
- when there is an unexpected increase or change in SARS-CoV-2 transmissibility and/or virulence
- when there is suspicion of a change in the performance of diagnostic (antibody, antigen, molecular assays) methods or therapies
- in risk settings, such as where there is close human–animal interaction with a large number of animals that are susceptible to SARS-CoV-2 infection, or where there are immunocompromised patients with prolonged shedding, especially when receiving antibody therapy against SARS-CoV-2.

For the reactive sequencing stream in the Irish national surveillance programme, there is a need to prioritise limited sequencing capacity to those activities with the highest clinical and/or public health potential. In the first instance, these will be restricted to investigations of complex outbreaks, suspected incidents of vaccine escape or anti-viral resistance, and unexpected changes in virulence and/or transmissibility.

WGS may provide a better understanding of transmission and other factors in complex outbreak investigations including: nosocomial outbreaks, travel-related outbreaks, Irish Traveller cluster links, possible super-spreader events, possibly linked outbreaks, and other complex healthcare setting outbreaks. Complex outbreak investigations may be particularly resource intensive, given that extensive genomic sampling of most identified patients in the epidemiological cluster of interest is required, as well as samples that are not part of the cluster being investigated. Samples from outside the cluster are important to support the hypothesis that cluster samples are epidemiologically linked more closely to each other than to other community infections. A single complex outbreak investigation could require over 100 sequences, and so sequencing should be reserved for outbreaks that would benefit the most from WGS. The Programme may be limited to a very small number of complex outbreak investigations per week, depending on cluster size.

**Investigation of potential vaccine escape** (i.e. SARS-CoV-2 infections in fully immunised individuals), **reinfections or anti-viral resistance** will become more important as COVID-19 vaccination programmes are implemented across the world, and more therapies for SARS-CoV-2 become available. When suspected cases arise it is important that these are fully investigated through WGS.

**Unexpected changes in virulence or transmissibility** may be identified through Stream 1 national surveillance, or may be detected from other surveillance or epidemiological investigations. Examples include unusual epidemiological patterns such as individuals or populations with higher attack-rate, increased disease severity or disease transmissibility, or a change in the age-related incidence. Consideration should be given to sequencing a subset of samples from an area with unusually high incidence. Given the potential for new variants of concern to become dominant rapidly within a population, it is critical that any such cases and clusters are urgently investigated using WGS.
Other situations (e.g., human-animal transmission, immunocompromised individuals undergoing antibody therapy, change in diagnostic performance), should be evaluated on a case-by-case basis as to whether to prioritise for WGS, given the current constraints on the system. These objectives may become more important in the medium to longer term (> 3 months). For example, WGS at the human-animal interface could become important if a new mutation of concern in animals with potential for transmission to humans was identified.

Requests for Stream 2 sequencing should be made by the regional Departments of Public Health to the HPSC. The HPSC in conjunction with the NVRL and regional Departments of Public Health should outline the request process for Stream 2 Reactive Sequencing, and develop prioritisation criteria for sequencing, especially regarding outbreaks and clusters. The HPSC should develop guidance on sampling of outbreaks for the purpose of sequencing to aid public health teams to gain the most benefit from phylodynamic analysis.

**Capacity requirements**

The overall aim is to sequence approximately 10% of cases. In the short term (i.e., three month timeframe) the National Programme will aim to have capacity to sequence up to 1,200-1,500 samples per week, with the potential to increase to 5,000 per week with surge capacity. Of the 1,200-1,500 samples to be sequenced weekly, 70-80% capacity should be reserved for Stream 1, with 20-30% capacity reserved for Stream 2. Based on current costs, at a run rate of 1,200 samples per week, sequencing costs alone would equate to an annual budget of approximately €9 million. The cost per sequence is likely to change over time given the potential for efficiencies including the use of higher throughput platforms and technological advances.

To meet this capacity need in the immediate term, the National Programme should engage with all available, publicly and privately funded laboratories currently conducting SARS-CoV-2 sequencing. The NVRL as lead diagnostic partner should lead on any outsourcing with contractual arrangements subject to approval by the HSE to ensure that running costs of increased sequencing activity can be met. The NVRL must be satisfied that any other laboratory conducting WGS on behalf of the National Programme has adequate quality management systems in place, and that data protection and other legal and ethical considerations are adhered to. This engagement and sub-contracting will be conducted in accordance with usual governance structures and in compliance with HSE procurement protocols, with particular consideration given to quality, efficiency, capacity, expertise, turn-around time and cost in selecting sub-contracted laboratories.

The national surveillance and WGS strategy developed for implementation in the medium- to longer- term (i.e., after three months) will determine the long term requirements for additional resources for diagnostics, bio-informatics and interpretation by public health expertise and for associated budget.

**Data flow**

Given that infectious disease surveillance falls under its remit, the HPSC will take the lead in the governance of the National Programme. As noted above, the NVRL will act as the primary site for coordination of sample preparation, sequencing and analysis, with outsourcing of samples, to other laboratories to meet immediate capacity requirements. The Clinical Director of the NVRL will be responsible for reporting sequence results to the HPSC,
in accordance with the core variables as specified by the HPSC. A data flow process will be agreed between the NVRL and the HPSC, building on existing processes.

The IEMAG will work with the HPSC to develop and provide the sampling frame for national community surveillance to the HPSC, based on up-to-date epidemiological data, as well as model the impact of variants on transmission, hospitalisations, ICU admissions, mortality and other disease outcomes, based on the surveillance data. Regular information flow will be necessary between the IEMAG, the NVRL and the HPSC.

The regional Departments of Public Health will liaise directly with the HPSC and the NVRL regarding Stream 2 sequencing and the results of subsequent investigations.

Sequencing results will be provided to HPSC to upload to the European surveillance system (TESSy) and to provide relevant sequencing information to other EU member states via the Early Warning and Response System (EWRS) alert system. The HPSC will also utilise sequencing results to provide surveillance reports to Department of Health and IEMAG. There will be an agreed process between HPSC, regional Departments of Public Health and the NVRL regarding the provision of sequencing results to Departments of Public Health, relevant clinicians and the affected cases of COVID-19. The NVRL and other sequencing labs, where appropriate, with the assistance of the HPSC, will continue to upload the relevant anonymised genomic data and linked metadata to international servers such as GISAID.

As recommended by the WHO, the National Programme will share anonymised meta-data along with SARS-CoV-2 genomic data, with international servers (e.g. GISAID) and warning systems (e.g. EWRS and TESSy), to maximise the utility of genomic sequencing. The rapid sharing of pathogen genome sequence data, both nationally and internationally, together with the relevant anonymised epidemiological and clinical metadata will maximise the impact of genomic sequencing in the public health response. Such data, generated during an outbreak, should be shared with the national and global community as rapidly as possible, to ensure maximum usefulness in improving public health.\(^3\)

It is important that WGS processes are underpinned by sound legal and ethical frameworks, ensuring that the data required to inform the national public health response are reported to the relevant national surveillance bodies, while safeguarding patient confidentiality and minimising patient harm. Human genomic sequences should be removed from the viral data at the earliest possible stage, unless ethical approval and explicit patient consent to process human genetic data have been obtained.\(^3\)

**Governance**

The governance structures for this national surveillance programme should be determined by the National Clinical Director of Health Protection and HPSC in line with their usual processes for operating communicable disease surveillance programmes. Consideration should be given to leveraging all relevant expertise which may include surveillance, public health, laboratory sequencing, bioinformatics and infectious diseases to inform the operational and governance processes.
# Appendix 1: Membership of the Working Group

<table>
<thead>
<tr>
<th>Chair</th>
<th>Director of Health Technology Assessment (HTA) &amp; Deputy Chief Executive Officer, HIQA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Working group members</strong></td>
<td></td>
</tr>
<tr>
<td>Dr John Cuddihy</td>
<td>Interim Director, Health Service Executive (HSE)- Health Protection Surveillance Centre (HPSC)</td>
</tr>
<tr>
<td>Dr Lorraine Doherty</td>
<td>National Clinical Director Health Protection, HSE-HPSC</td>
</tr>
<tr>
<td>Dr Patricia Harrington</td>
<td>Deputy Director, HTA Directorate, HIQA</td>
</tr>
<tr>
<td>Prof. Mary Horgan</td>
<td>President of the Royal College of Physicians Ireland &amp; Consultant in Infectious Diseases, Cork University Hospital.</td>
</tr>
<tr>
<td>Dr Cillian de Gascun</td>
<td>Consultant Virologist &amp; Director of the National Virus Reference Laboratory, University College Dublin</td>
</tr>
<tr>
<td>Prof. Paddy Mallon</td>
<td>UCD Full Professor of Microbial Diseases and Consultant in Infectious Diseases, St Vincent's University Hospital</td>
</tr>
<tr>
<td>Dr Lois O’Connor</td>
<td>Specialist in Public Health Medicine, HSE-HPSC</td>
</tr>
<tr>
<td>Michael O’Leary</td>
<td>Principal Officer, Department of Health</td>
</tr>
<tr>
<td>Dr Darina O’Flanagan</td>
<td>Special Advisor to the National Public Health Emergency Team (NPHET)</td>
</tr>
<tr>
<td>Dr Matt Robinson</td>
<td>Specialist Registrar in Public Health Medicine, HSE-HPSC</td>
</tr>
<tr>
<td>Dr Kieran Walsh</td>
<td>Senior HTA Analyst, HTA Directorate, HIQA</td>
</tr>
</tbody>
</table>
Appendix 2: Terms of reference of the Working Group

The terms of reference were to:

1. Provide recommendations on the appropriate governance structure of the programme and the lead organisation.
2. State the aim or aims of the national surveillance programme and how the performance of the programme in achieving its aims should be measured.
3. Review any relevant guidance documents from international organisations e.g. WHO, ECDC, or Public Health bodies in other jurisdictions.
4. Advise on the role of laboratory accreditation in the national surveillance programme.
5. Describe how the data from any new programme will be incorporated into existing notification and reporting processes, e.g. HPSC, CIDR, DoH, regional DPHs, hospitals, GPs, and other relevant stakeholders.
6. Identify the settings in which surveillance should take place e.g. primary care, community, acute hospitals, residential care, or other, and advise on the minimum reporting requirements from each setting.
7. Clarify the role of the patient, i.e. should the programme have the ability to feed data back to the patient, in case SARS-CoV-2 lineage becomes clinically relevant in the future e.g. in a case of potential antiviral therapy resistance, or indeed severe disease.
8. Determine what the national whole genome sequencing capacity (WGS) should be, and how this should be distributed between healthcare (medical diagnostic laboratories) and non-healthcare settings.
9. Consider the role for non-WGS approaches to SARS-CoV-2 surveillance.
10. Quantify the appropriate surge capacity for the national programme to allow for the prompt investigation of outbreaks of concern (as seen recently in the UK).
11. Ascertain the proportion and demographic composition of samples that needs to be sequenced on a weekly/monthly basis for SARS-CoV-2 surveillance to ensure that the data are nationally representative: this should be informed by input from IEMAG.
Appendix 3: Background information

Background

Genomic sequencing is defined as "a laboratory method that is used to determine the entire genetic makeup of a specific organism or cell type,"\(^{(8)}\) and may involve either partial or whole genome sequencing (WGS).\(^{(1, 3)}\) WGS can be used to investigate viral pathogen genomes, understand outbreak transmission dynamics and identify mutations that may have an impact on countermeasures.\(^{(1, 3)}\) Disease surveillance is defined as "an information-based activity involving the collection, analysis and interpretation of large volumes of data originating from a variety of sources."\(^{(9)}\) Given the wealth of data obtained by WGS, it has the potential to enhance disease surveillance, guide public health action and inform vaccination strategy.\(^{(1, 3, 10)}\)

In relation to SARS-CoV-2, WGS has several important applications. These include:

- understanding the emergence of SARS-CoV-2 (e.g., identifying the causative agent of COVID-19 and the zoonotic origin(s) of SARS-CoV-2)
- understanding the biology of SARS-CoV-2 (e.g., understanding host receptor usage and virus evolution)
- improving diagnostics and therapeutics (e.g., improving molecular diagnostics, supporting vaccine and anti-viral design, and identifying antiviral resistance and vaccine-escape mutations)
- investigating virus transmission (e.g., outbreak investigation and identifying importation events)
- inferring epidemiological parameters (e.g., reproduction number).\(^{(3)}\)

Since the recent emergence of variants of concern - including VOC 202012/01 or B.1.1.7 lineage first identified in the United Kingdom (UK) and 501.V2 or B.1.351 lineage first identified in South Africa in late 2020,\(^{(11)}\) and more recently P.1 or B.1.1.248 lineage first identified in travellers returning from Brazil in January 2021\(^{(12)}\) - there has been an increased emphasis on the important role of WGS and surveillance to detect and monitor the incidence of these new variants, and inform public health action.\(^{(11)}\) While viruses constantly mutate, these recently identified variants are of particular concern due to emerging evidence of increased transmissibility compared with previously circulating strains.\(^{(13)}\) There is also some concern that the UK variant of concern may be associated with more severe disease progression,\(^{(14)}\) and that the South African variant may be less susceptible to neutralising antibodies (NAbs) targeting the original SARS-CoV-2 spike protein.\(^{(15-17)}\)

In its latest risk assessment published on 21 January 2021, the European Centre for Disease Prevention and Control (ECDC) assessed the probability of the introduction and community spread of these variants of concern in the European Union/European Economic Area (EU/EEA) as very high due to their apparent increased transmissibility.\(^{(18)}\) Such increased transmissibility is likely to lead to an increased number of infections. This, in turn, is likely to lead to an increase in the total number of hospitalisations and deaths across all age-groups, but particularly for those in older age groups or with co-morbidities. The possibility that variants of concern may lead to more severe disease and or reduced susceptibility to vaccine-elicited NAbs would also have implications for public health and vaccination strategies. The ECDC recommends that member states need to increase the level of surveillance and WGS in order to monitor the spread of these variants. The European Commission has urged member states to ramp up national genomic surveillance to 5-10% of
all positive test results, and the Commission also underlined the commitment at a European level to increase sequencing capacity across the bloc, by offering member states capacity for WGS through ECDC labs.\(^{(5)}\)

The main public health objectives of SARS CoV-2 sequencing are as follows:\(^{(1)}\)

- Early detection and characterisation of emerging variant viruses to define if they are of particular concern; early detection of variant viruses is key to the implementation of public health response measures in a timely manner in order to reduce the impact of the variant.
- Assessing the impact of genetic and antigenic variant viruses for the pandemic and monitoring them over time to guide public health action.

**Conduct of WGS for SARS-CoV-2**

WGS is conducted on the residual material from samples used for PCR analysis. Prior to WGS the sample must undergo a number of preparatory steps that are labour intensive, including viral RNA extraction, cDNA generation and sequence library preparation. Multiple samples are sequenced simultaneously over a 48 to 72 hour period. Sequences generated are then analysed by bio-informatics to identify the viral clade, lineage and mutations of interest. Sequences linked with patient epidemiological and clinical metadata (i.e., anonymised data including sex, age, county of residence) are uploaded to international databases such as the Global Initiative on Sharing All Influenza Data (GISAID) to enable international surveillance and research. New variants are also quickly identified if they differ significantly in sequence from existing variants. All changes in amino acid sequence are also revealed in the sequence data of the virus variant that has been identified from each patient. This information is crucial to quickly identify possible increased or decreased virulence, pathogenicity or vaccine resistance. Information on viral clade, lineage and mutations of interest comprise the data to inform national and targeted public health actions and potentially vaccination strategy. WGS is near real time analysis. Currently in Ireland the process from sampling to results to inform public health action takes between seven and 14 days depending on the setting.

In order to undertake WGS, there is a technical requirement for a sufficiently high viral load in a sample, as well as sufficient sample volume.\(^{(3)}\) This eliminates a large proportion of all potential samples (estimated to be approximately 40-50%). Samples need to be collected as early as possible in the course of the infection (ideally within 5-7 days of symptom onset) given that SARS-CoV-2 viral load peaks early and declines relatively quickly after the first week.\(^{(19)}\) Viable virus that is required for sequencing may not be obtainable after 9-10 days of symptoms; however, immunocompromised individuals may shed viable virus for longer.\(^{(20)}\)

It is important that WGS processes are underpinned by sound legal and ethical frameworks, ensuring that the data required to inform the national public health response are reported to the relevant national surveillance bodies, while safeguarding patient confidentiality and minimising patient harm. Human genomic sequences should be removed from the viral data at the earliest possible stage, unless ethical approval and explicit patient consent to process human genetic data have been obtained.\(^{(3)}\)
Current SARS-CoV-2 surveillance and WGS operations in Ireland

Notification requirements

All medical practitioners, including clinical directors of diagnostic laboratories, are required to notify the Medical Officer of Health (MOH) or Director of Public Health (DPH) of certain diseases, including COVID-19. In turn, the MOH is required to notify the Health Protection Surveillance Centre (HPSC) of all cases of notifiable infectious diseases. The list of diseases (and their respective causative pathogens) that are notifiable is contained in the Infectious Diseases Regulations 1981 and subsequent amendments. Notification is in accordance with the case definitions. Case definitions used in Ireland are predominantly based on European Commission (EC) case definitions. The HPSC is responsible for developing, maintaining, updating and circulating the case definitions for each notifiable disease. The most recent amendment to the Regulations is the Infectious Diseases (Amendment) Regulations 2020 (S.I. No. 53 of 2020). Notification is in accordance with the core variables on the surveillance system, Computerised Infectious Disease Reporting (CIDR).

The core variables are specified by the HPSC; additional non-mandatory fields can be added by the HPSC to inform surveillance and public health response and control. The HPSC also has the authority to declare any variant to be of ‘public health concern’ and communicates regularly with both public and private laboratories in relation to their legal obligations in respect of notification of infectious diseases, including variants of concern.

Under the Infectious Diseases regulations, the legal responsibility lies with clinical directors of diagnostic laboratories to report data in relation to notifiable diseases, including sequencing data of variants, to the HPSC as soon as it is identified in the laboratory. Notification to the MOH or DPH is a legal obligation and is not in contravention of data protection legislation. The MOH or DPH is required to treat records of infectious disease notifications in a confidential manner. Individual patient consent is not required to allow diagnostic data and patient clinical metadata to be used for national surveillance purposes.

Laboratory notifications are made electronically through CIDR, which is a web-based system that holds all data in a single national information repository. The system was developed to manage the surveillance and control of infectious diseases in Ireland. Information from laboratories is entered manually or uploaded electronically into CIDR by laboratory scientists/microbiologists. This information is then linked to clinical and epidemiological information provided by public health professionals. There is currently no framework by which WGS information is linked with individuals on CIDR, and so WGS is currently not routinely utilised for enhanced surveillance purposes in Ireland. HPSC is in the process of adding WGS fields to CIDR, so that data on WGS can be included in the electronic laboratory notification of SARS-CoV-2, building on the existing model and pathways developed between HPSC and notifying laboratories.

WGS activity in Ireland

Currently, there are two entities conducting the majority of WGS of SARS-CoV-2 samples in Ireland. These are the National Virus Reference Laboratory (NVRL) and the Irish Coronavirus Sequencing Consortium (ICSC). It is important to acknowledge that there has been significant collaboration between the ICSC and NVRL since the onset of the COVID-19 pandemic. NVRL is a partner in ICSC. NVRL has undertaken sequencing activity to inform
the operational public health response. NVRL has also undertaken sequencing on behalf of the consortium for its research activities funded by Science Foundation Ireland.

Since February 2020, a total of 2,212 Irish sequences have been uploaded to the international server GISAID, equating to about 1.2% of all COVID-19 cases in Ireland.(23)

The NVRL sequencing activity to inform the public health operational response has included 1,071 SARS-CoV-2 specimens to date. Of these 1071 specimens, 873 were processed completely (extracted, sequenced, and analysed) at the NVRL; the remaining 198 were sequenced at UCD CFS (Centre for Food Safety) on behalf of the NVRL, with RNA extraction and data analysis performed at the NVRL. In addition to this 'NVRL' work, the NVRL has sequenced 186 specimens for the ICSC (these samples would have been initially extracted in CEPHR (the Centre for Experimental Pathogen Host Research in University College Dublin). (24)

In total, the ICSC, has sequenced more than 1,000 SARS-CoV-2 samples to date. In total, 362 samples are being sequenced in collaboration with the All Ireland Infectious Diseases (AAID) cohort. The ICSC has also sequenced other samples from clinical and community settings where consent or a waiver from consent was obtained.

There are other laboratories in Ireland (both publicly- and privately funded), outside of the NVRL and the ICSC, who are currently conducting, or who have the capacity to conduct WGS of SARS-CoV-2 samples. National capacity includes a number of hospital laboratories, academic laboratories and private entities. There is also the option to avail of international laboratories, including the German private provider who currently augments national PCR diagnostic capacity as well as those run by the ECDC, to complement and supplement national WGS capacity.(5)

**National Virus Reference Laboratory**

Established in 1963, the NVRL is the existing lead laboratory partner for HPSC for its national virus surveillance programmes, such as influenza surveillance. The NVRL is also the WHO accredited national laboratory for measles, rubella, polio and non-polio enterovirus surveillance. During each influenza season, samples are provided to the NVRL from 60 sentinel GP practices in the existing primary care influenza-like illness (ILI) sentinel surveillance system. Genetic and antigenic characterisation of influenza samples informs international identification of dominant circulating strains thereby informing selection of strain-specific influenza vaccines for national vaccination programmes. Existing hospital surveillance for emerging influenza strains associated with more virulent disease, is conducted via characterisation of samples from 5% of hospitalised patients with more severe presentation of disease. These data are forwarded to the NVRL by the hospital diagnostic laboratories at request of HPSC.

In establishing a national SARS-CoV-2 WGS and surveillance programme, there are opportunities to build on the current surveillance structure and already established linkages for sample and data transfer between HPSC, departments of public health, NVRL, sentinel GP practices and the hospital laboratories

The NVRL has oversight of the majority of national PCR-based diagnostic activity for SARS-CoV-2, which is carried out by a network of UCD NVRL laboratories located at UCD Belfield, Enfer Laboratories in Sallins and an NVRL Satellite laboratory at Backweston Laboratory.
Campus in Celbridge. Through this network, NVRL has access to samples from all settings that are being processed as part of the routine SARS-CoV-2 community diagnostic pathway.

Following an earlier endorsement by NPHET to increase WGS capacity, HSE funding has already been approved to expand NVRL infrastructure adding capacity for a further 250 to 500 samples per week. This additional capacity should be in place within approximately a 2-3 month time frame.

In the interim, the NVRL has recently begun to increase WGS capacity utilising current infrastructure and leveraging academic laboratory support to generate about 180 whole genomes per week (in conjunction with the Centre for Food Safety (CFS) in UCD). Since the summer, the NVRL has - using the Oxford Nanopore MinION - been sequencing 47 samples (the standard capacity for a single flow cell) every 1-2 weeks (weekly in recent months) as a baseline. With the increase in demand in recent weeks and months, the NVRL has engaged with Prof. Fanning in the CFS in UCD, who has also generated 47 sequences per week. These samples are extracted and the data analysed at the NVRL. To increase capacity further - while awaiting completion of the procurement process for additional equipment as mentioned above – the NVRL is now doubling the number of samples on a single flow cell to 94. Doubling the number of samples will reduce the number of reads generated per sample; however, the amount of reads generated for 94 samples will still be more than sufficient to determine the lineage/variant of the SARS-CoV-2 in the specimen. With both the NVRL and CFS doubling the number of samples of a single flow, this should yield a weekly capacity of around 180 samples (allowing for some failures). (24)

Up until recently, the sampling approach used for WGS in Ireland has not been based on any formal nationally representative framework. Since the beginning of the pandemic, in addition to a random selection of positive specimens from the community, the NVRL has also processed samples in response to requests from public health departments for WGS with the aim of increasing the understanding of outbreaks with unusual transmission patterns. More recently, the NVRL has sequenced a more targeted selection of samples from individuals who have tested positive for SARS-CoV-2 having travelled to Ireland from countries in which the variants of concern are thought to have originated.

The NVRL has recently received a sampling frame devised by the Irish Epidemiological Modelling Advisory Group (IEMAG). This framework is based on a standard PCR plate of 96 samples and comprises a nationally representative geographic and demographic distribution. The sampling frame was initially created to monitor the distribution of SARS-CoV-2 lineage B.1.1.7. However, as the NVRL weekly WGS capacity increases (as described above), this sampling frame can also be used for WGS. IEMAG has offered to increase the size of the sampling frame in line with available WGS capacity. Currently samples are not identified for WGS based on clinical disease progression.

With regards to non-genomic surveillance for variants of concern, the NVRL is increasing diagnostic capacity at its Backweston facility. Both the NVRL facility in UCD and the Backweston facility are operating the ThermoFisher TaqPath assay. This facilitates real-time surveillance for the UK variant as part of the routine diagnostic pathway, by screening for S gene drop-out (due to its 69/70 deletion) and triage of these samples for confirmatory WGS, as occurs in other countries. The NVRL has also introduced a new allele-specific PCR test to detect the N501Y amino acid change which is one of the key mutations in the recently described novel South African SARS-CoV-2 variant. Samples expressing the N501Y amino
acid change are triaged for confirmatory WGS. Given that the majority of mutations causing concern occur in the spike protein, Sanger sequencing - which involves sequencing of individual fragments and is generally widely available and easy to use — is another potential option for surveillance for variants of concern. Non-WGS surveillance that facilitates early identification and potential triage of samples with a higher probability of being associated with a variant of concern may provide a more cost-effective and efficient screening process than WGS alone.

**Irish Coronavirus Sequencing Consortium (ICSC).**

The ICSC was established in early 2020 as part of a Science Foundation Ireland (SFI)-funded COVID-19 Rapid Response grant to provide capability for large scale sequencing of SARS-CoV-2 samples to map the evolution of the virus in Ireland. The ICSC is led by Prof Paul Cotter from Teagasc.¹ Viral RNA extraction and further preparation is undertaken by CEPHR in UCD.² Sequencing is undertaken at two Teagasc laboratories and a number of other publicly funded laboratories in addition to the NVRL. Commercial laboratories are also partners in the consortium. Bioinformatic analysis for ICSC is undertaken by Teagasc.

ICSC activity to date is limited to research with appropriate ethical approval and patient consent. ICSC has primarily sequenced hospital samples with ethical approval and relevant consent for analysis obtained through the All Ireland Infectious Diseases Cohort (AIID Cohort) study. The AIID study led by Prof Paddy Mallon, UCD, is a prospective cohort study of patients with COVID-19 in hospitals in 11 centres nationally.²⁵

For the AIID cohort study, patients in the 11 study hospitals that meet the inclusion criteria and provide informed consent are enrolled in the study. Samples from the study participants are sent to the study labs to be sequenced. Given the prospective nature of the AIID cohort, samples obtained from patients and sequenced can be linked to the maximum disease severity that the patient experiences, which may occur later in the hospitalisation episode, for example, intensive care unit (ICU) admission or death.

At present ICSC weekly capacity for WGS, excluding NVRL, based on Oxford nanopore technology is 200 sequences with ability to increase quickly to 300 sequences weekly. Increasing available capacity for WGS at the Teagasc laboratories beyond 300 samples per week, involving use of the Illumina sequencing platforms at their disposal, is already underway in collaboration with collaborative SARS-CoV-2 sequencing laboratories in the UK. This additional infrastructure could come on line within a matter of weeks.

---

¹ Teagasc – the Agriculture and Food Development Authority – is the national body providing integrated research, advisory and training services to the agriculture and food industry and rural communities.
² CEPHR – The Centre for Experimental Pathogen Host Research, in University College Dublin is led by Prof. Paddy Mallon.
References


