Alberta COVID Variants Surveillance Plan (version 4, January 26, 2020)

Genomics surveillance strategy development: Graham Tipples, Nathan Zelyas, Kanti Pabbaraju and Matthew Croxen
Reporting: Stephanie Murphy and Emily Buss

Updates from previous version:
- Daily variant screening (300) and weekly virus sequencing (400) projections added.
- Update of annexes 1 and 2 – information of SARS-CoV-2 lineages and variants of concern
- Addition of annex 4.

Background

The causative agent of COVID-19 is SARS-CoV-2. COVID-19 virus genomics is the study of the viral RNA genome sequence (29.9 kilobases (kb)). As with many viruses, replication of SARS-CoV-2 is error prone and the virus evolves over time. Over the course of the pandemic since the initial emergence of this virus, numerous lineages of SARS-CoV-2 have evolved. These lineages may be widespread or limited to particular geographic regions. In addition, the changes in the nucleotides that make up the genome sequence may or may not result in biological changes to the virus. A change in a particular nucleotide is often referred to as a single nucleotide polymorphism (SNP). The accumulation of numerous SNPs over time is what results in the evolution into a distinct lineage. See annex 1.

The recent reporting of several variants of concern (e.g. UK, South Africa and Brazilian variants) are of interest because of the presence of particular mutations or SNPs putatively associated with important biologic activity. See annex 2 for details on the mutations of potential biological significance and for characteristic mutations that define the emerging variants of concern.

The UK variant was discovered through the combination of ongoing genomic analysis together with monitoring the spread of COVID-19 referred to as molecular epidemiology. The increased spread in terms of numbers of cases seen in the UK has been thought to be at least partially attributable to the changes in the virus, specifically mutations in the spike protein binding to the host cell, resulting in more efficient binding and enhanced transmissibility.

There will inevitably be more mutations that emerge of biologic significance. There are two important components to the surveillance for the emergence of variants of concern.

First, there needs to be the robust background context and knowledge for the evolution and distribution of SARS-CoV-2 lineages in Alberta, Canada and internationally. All throughout the pandemic APL-Public Health Lab (ProvLab) has been working as part of a national surveillance initiative called the Canadian COVID-19 Genomics Network (CanCOGen). The anonymized virus sequence data (whole virus genome sequencing) goes into this centralized database so we have the background context to assess the emergence of new variants. The full genome sequencing is essential but not amenable to high volumes of testing nor necessarily timely.

Second, once a variant of concern or particular mutations of biological significance are reported, then more targeted and timely screening tests can be developed to look specifically for certain variants of concern. There can be various approaches to these targeted screening tests. One approach is the use of a commercial test that may have one of their diagnostic targets (69/70 deletion) affected thus giving an indication of the presence of the mutation (this is exactly the case with the Thermofisher TaqPath assay). Other approaches are lab-developed SNP PCR assays which can be used to target SNPs of interest. ProvLab is using both of these approaches.

The evolution of SARS-CoV-2 and the detection of the emergence of variants of concern are important to monitor for strains that can contribute to:

- enhanced transmissibility (i.e. contribute to more rapid spread)
- increased virulence (i.e. cause more severe illness)
• diagnostic escape mutants (i.e. mutant strain may escape detection by particular tests)
• escape from prior immunity conferred by either previous COVID-19 infection or COVID-19 immunization

**Variant Testing strategy**

The strategy for detection of variants will consist of continued genomic surveillance as well as targeted screening. Targeted screening will enable identification of known SNPs while the genomic surveillance will provide the ongoing capacity to detect new lineages and variants. Our Alberta strategy for SARS-CoV-2 genomics surveillance and screening for variants aligns with the CanCOGen strategy and priorities for the national strategy.

• random sampling: positive samples that have been submitted to the public health labs along with some submitted to other labs
• targeted groups:
  o international travelers
  o suspected reinfection
  o post-vaccine infection
  o severe disease in healthy/young
  o pediatric cases
  o outbreaks or superspreading events

**Methods:**

• The workflow for identification of samples of interest for variant screening and ongoing surveillance and reporting of results is shown in annex 3.
• Screening tests
  o relatively fast TAT, less information (can only look for one SNP at a time)
  o these assays will generally not determine the specific lineage or variant by themselves; they are quick tools to detect SNPs of biological interest and to flag potential variants of concern
  o TAT target is 24 h after we identify that a sample requires screening
  o Projected daily screening test volume is approximately 300 variant tests daily.
  o The screening assay detects the 501 mutation and the 69/70 deletion.
• Sanger sequence of regions of the S gene
  o less fast TAT
  o provides more information about multiple nucleotide positions
• Next generation sequencing (NGS)
  o slowest TAT
  o provides most information about the isolate, including lineage
  o TAT target is 5 days for NGS after a sample is identified for priority NGS when a potential variant of concern is flagged by a screening assay
  o TAT will be longer for samples that are of lower priority (e.g., investigations for reinfection)
  o Projected weekly whole virus sequencing is approximately 400 sequences weekly.

**Reporting:**

• Work is ongoing to get these results into our lab information system (LIS) for streamlined reporting.
• In addition, work is underway for optimal reporting taking into account overall understanding of the evolving lineages, the emergence of new variants of concern and the detection of specific mutations of potential biological significance. It’s not as straightforward as a positive or negative PCR result!
• The goal is to have a regular report summarizing:
  o Number of samples screened for the variants of concern
  o Number of positive screening results
  o Number of samples with sequencing completed
  o Number of variants of concern identified
  o Demographic information for samples undergoing screening and sequencing
Proportion of samples genotyped:

- Using the screening tests, we are hoping to test all positive samples from the targeted groups (including a subset of samples from the outbreaks superspreading events as not all would need to be sequenced by NGS)
- All positive samples detected at the public health laboratories (which comprise ~1/3rd of all positive samples in the province) will be subjected to testing via the screening assays
- If case numbers continue to decrease, we will be able to increase the proportion of random positive samples that are subjected to the screening tests
- We are aiming to run ~300/day on the screening tests once they are up and running
- Sanger sequencing and NGS will be used to characterize samples that are positive using the screening tests along with a random sampling of positives detected throughout the province submitted to the public health laboratories. Projected sequencing capacity is 400 weekly.
- NGS would be used for suspected reinfections (to compare the first and subsequent strains) and post-vaccine infections (as escape mutants could arise that may not be detected by the screening assays).

Caveats:

- Rapid testing may or may not have a second sample available from which to do the variant and/or genomic analysis.
- Limited freezer and fridge space for retention of samples may limit retrospective analysis of investigations (e.g. reinfections).
- Large number of labs and decentralized testing using rapid tests may limit ability for full representation of the province for sampling.
- Not all positive samples will have a high enough viral load to perform reliable genotyping.

Annexes

1. SARS-CoV-2 lineages
2. Variants of concern: defining/characteristic mutations
3. SARS-CoV-2 variant screening workflow
4. Traveler identification workflow
Annex 1: SARS-CoV-2 lineages

(a) Global SARS-CoV-2 lineages throughout the pandemic

(b) Alberta SARS-CoV-2 lineages from September 2020 through January 2021
(c) Relative proportion of lineage detected in Alberta from September 2020 through January 2021

(d) Number of lineages detected in Alberta from September 2020 through January 2021
Annex 2: Variants of concern: defining/characteristic mutations

(a) Key mutations in the spike protein of potential biological significance:

<table>
<thead>
<tr>
<th>Mutation</th>
<th>Putative biological significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>N501Y</td>
<td>a mutation in the receptor binding domain, involved in viral binding to human cells</td>
</tr>
<tr>
<td>69/70 deletion</td>
<td>a 6 nucleotide deletion the spike gene resulting in a spike protein change thought to be involved in the virus binding; may be involved in reduced recognition of antibodies</td>
</tr>
<tr>
<td>P681H</td>
<td>adjacent to the furin cleavage site (necessary for processing to S1 and S2), a site which may promote increased infectivity</td>
</tr>
<tr>
<td>E484K</td>
<td>a mutation that may reduce neutralization of convalescent serum, found in reinfection case in Brazil</td>
</tr>
<tr>
<td>K417</td>
<td>a mutation in the receptor binding domain, involved in viral binding to human cells</td>
</tr>
</tbody>
</table>

(b) Variants of concern (VOC) and defining mutations in the spike protein

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Key mutations of potential significance</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>UK</td>
<td>N501Y, 69/70 deletion, P681H</td>
<td>Reported in 55 countries, travelers and community spread; Sporadic travel related cases in Alberta</td>
</tr>
<tr>
<td>South Africa</td>
<td>N501Y, K417N, E484K</td>
<td>Reported in 22 countries, mostly related to travel</td>
</tr>
<tr>
<td>Brazil / Japan</td>
<td>N501Y, K417T, E484K</td>
<td>Reported in two countries: Brazil (community), Japan (travel)</td>
</tr>
</tbody>
</table>

(c) Emerging variants of concern

<table>
<thead>
<tr>
<th>Originally described</th>
<th>Lineage</th>
<th>Key mutations of potential significance</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brazilian variant</td>
<td>P.2</td>
<td>E484K</td>
<td>Related to P.1, described in Brazil</td>
</tr>
<tr>
<td>Nigerian variant</td>
<td>B.1.1.238</td>
<td>P681H</td>
<td>Reported in media, based on two genomes</td>
</tr>
<tr>
<td>US variant</td>
<td>B.1.2</td>
<td></td>
<td>Widespread in North America; spike mutation of unknown relevance</td>
</tr>
</tbody>
</table>
Annex 3: SARS-CoV-2 variant screening workflow
Annex 4: Traveler identification workflow