



# CanCOGeN

## UPDATE July 2023

### Overview

Funded with \$40 million from the federal government, Genome Canada launched the [Canadian COVID-19 Genomics Network \(CanCOGeN\)](#) in April 2020. CanCOGeN was a Genome Canada-led network of Canadian federal, provincial and regional public health authorities and their healthcare partners, academia, industry, hospitals, research institutes, large-scale sequencing centres and small sequencing facilities of provincial health labs across the country. The initiative consisted of two components, [VirusSeq](#) and [HostSeq](#), established to coordinate a pan-Canadian, cross-agency network for large-scale SARS-CoV-2 and human host sequencing to track viral origin, spread and evolution, characterize the role of human genetics in COVID-19 disease and inform time-sensitive critical public health and policy decision making relevant to health authorities across Canada during the pandemic. CanCOGeN has overcome some complex challenges and achieved several significant impacts, setting up the foundation for national viral genomic surveillance in Canada. This short paper provides a high-level summary of key data-related challenges and impacts from CanCOGeN. Broader impacts and lessons learned from this cross-sectoral national initiative are captured in Genome Canada's [CanCOGeN newsletter](#) series as well as a comprehensive case study by the [Public Policy Forum](#).



### NUMBERS

**CanCOGeN VirusSeq: Originally tasked to sequence up to 150,000 SARS-CoV-2 genomes and has sequenced 500,000 viral genomes (as of June 2023)**

**CanCOGeN HostSeq: Goal to sequence up to 10,000 individuals exposed to, or affected by, the SARS-CoV-2 virus and has sequenced 11,300 host participants (as of June 2023)**



Credit: SFU Brinkman Lab, Patrick Taylor. Photo by Aleksa Nenadic

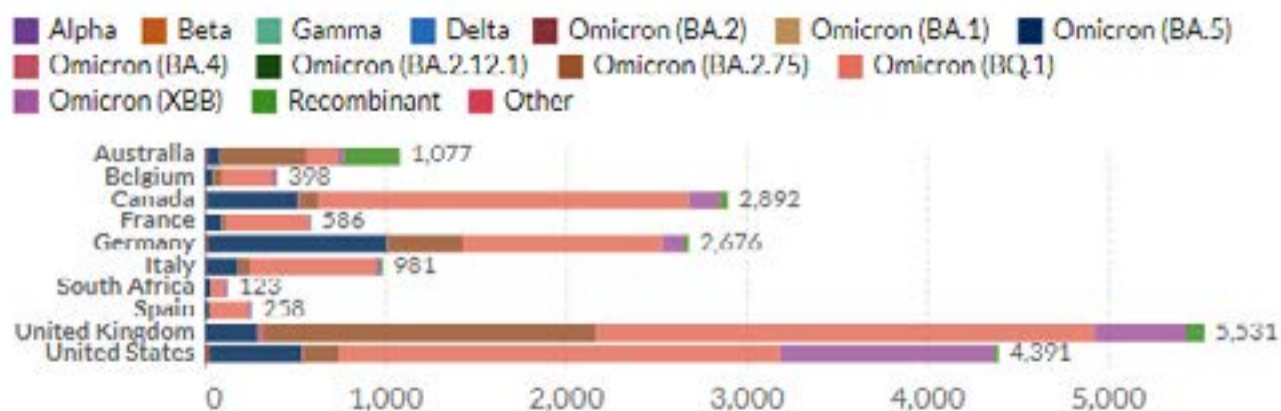
### Data sharing

Data sharing (within provinces, across provinces and internationally) proved to be the most significant challenge CanCOGeN had to overcome. A Genome Canada-led interim evaluation of CanCOGeN identified the critical need for clarity on data sharing requirements (for genomic and metadata) within the initiative. In other words, creating standards for sequencing; identifying minimum metadata associated with viral sequences; having agreements on early use of data and material sharing; and ensuring engagement of data sharing experts. CanCOGeN met these challenges by building robust governance structures, creating effective working groups, developing data sharing protocols and standards, and submitting sequencing and metadata information into public international databases like GISAID, the Global Initiative on Sharing Avian Influenza Data. Data sharing, first through GISAID and then through a CanCOGeN-led Canadian VirusSeq Data Portal, was critical to ensure Canada's alignment with global health measures aimed at understanding and controlling COVID-19.

## Data impacts

- Two critical metadata standards were developed by CanCOGeN: Contextual<sup>1</sup> metadata standards (used at the Public Health Agency of Canada's (PHAC) National Microbiology Laboratory (NML) and provincial health labs), and Minimum metadata standards (for data shared with national and international databases). Getting an agreement between provincial public health labs and PHAC on Contextual and Minimum Metadata was critical to advance Canadian data sharing initiatives. These new standards are now in use for new and emerging pathogen surveillance.
- CanCOGeN also established user standards to identify who is contributing to different components of genomic surveillance initiatives, so all contributions can be properly acknowledged, and data providers and data users can connect for coordination and collaboration.
- CanCOGeN launched Canada's first open genomic data sharing network in April 2021 to track SARS-CoV-2 variants in real time: the [Canadian VirusSeq Data Portal](#). This collaboration between academics led by Dr. Guillaume Bourque (and including Drs. Fiona Brinkman, William Hsiao, Lincoln Stein and Yann Joly) and Canadian genomics software company DNASTack was an important deliverable for CanCOGeN and helped advance data sharing initiatives in Canada.
- The collaboration with DNASTack for the Canadian VirusSeq Data Portal expanded the application of Viral AI, a tool to access, interrogate and analyze large scale national and international viral datasets. Viral AI was developed by DNASTack and a national consortium supported by Canada's Global Digital Supercluster.
- As a key pillar of its strategy, CanCOGeN also supported the development of the [HostSeq DataBank](#), the first large-scale Canadian cohort of individual whole genome sequences linked to detailed clinical information. HostSeq houses 17 studies across the country with additional host genome data being continually added. The Databank is open to scientists across Canada that are interested in using this data.
- The HostSeq DataBank will form the foundation for the Canadian Institutes of Health Research (CIHR) and Genome Canada-partnered [Canadian Human Genome Library](#) (CHuGL), a federated pan-Canadian framework that will enable access to, and analysis of, human genomes and associated health/environmental information, governed in a transparent, accountable and consistent way.
- The lessons from CanCOGeN's data sharing activities informed the work of the Pan-Canadian Health Data Strategy through membership on its expert advisory group, bringing knowledge, experience and solutions to the fore for health data sharing across the country.
- Lessons learned from CanCOGeN continue to improve genomic sequencing data quality and interoperability across Canada to support public health use. Genome Canada and the NML have collaborated to create a joint initiative, the Canadian Public Health Alliance for Research in Genomic Epidemiology and Surveillance (CHARGES). CHARGES will help coordinate future viral sequencing, and both organizations collaborate to maintain data sharing and the ongoing work of the Canadian VirusSeq Data Portal.
- With the new data capacity built in Canada, members from many different research institutes and public health agencies have joined the international Public Health Alliance for Genomic Epidemiology (PHA4GE), to improve the genomic data capacity, and public health bioinformatics tools and infrastructure interoperability, reproducibility, and portability across international jurisdictions.
- Since CanCOGeN's inception, Canada has moved from a lagging position in terms of global submission of data on variants, to a top three spot among countries submitting viral data (*Figure 1*).

**Figure 1. COVID-19 variants by country reported to GISAID international database January 15 and 29, 2023**



Source: GISAID, via CoVariants.org - Last updated 29 January 2023

OurWorldInData.org/coronavirus • CC BY

Note: Recently-discovered or actively-monitored variants may be overrepresented, as suspected cases of these variants are likely to be sequenced preferentially or faster than other cases.

<sup>1</sup> These data types include sample collection and processing information, exposures, symptoms, pre-existing conditions, vaccination, reinfection, sequencing, bioinformatics and diagnostic testing information, as well as variant information.

## Ecosystem impacts

- The genomic data generated through CanCOGeN became one of the strongest tools for pandemic surveillance—tracking the virus’ evolution and the spread of novel variants of concern. CanCOGeN’s investment has also strengthened genomic sequencing capacity, skills, data sharing and collaboration across the country, bolstering capacity to tackle other major public health challenges and future pandemics.
- Increased sequencing capacity for public health has been realized by CanCOGeN through both [soft infrastructure for genomic surveillance](#) (people—over 50 highly skilled FTEs and 16 trainees, skills, three new protocols, and [eight new software tools](#)) and hard sequencing infrastructure (sequencing technology and related reagents) leading to nine provinces sequencing in-province (compared with four at the onset of COVID-19). This new capacity supports provinces to better respond to local outbreaks of COVID-19 in the current pandemic, and to address future pathogen outbreaks with fast local public health solutions.
- Provincial public health labs became more closely aligned with genomics programs available through regional Genome Centres. This creates new relationships and gives the public health labs the opportunity to apply for new funding to advance genomics research facilitating the adoption of genomics findings and technologies by public health laboratories. The National Microbiology Laboratory is now successfully partnering on the Genome Canada [Genomic Applications Partnerships Program](#) (GAPP) program.
- Direct link of genomic sequencing to policy outcomes made by provincial and federal government. For example, the closure of Whistler, B.C., ski resort in March 2021 was a direct result of variant sequencing showing the prevalence of the Gamma variant specifically in the town. Also, as new variants have been identified in Canada that related to travel (e.g., initial Omicron identification in November 2021), the federal government instigated travel restrictions and new testing policies that were a direct response to genomic surveillance capacity across the country.
- Thanks to the lessons learned from CanCOGeN, Genome Canada provided evidence-informed advice and expertise to key federal COVID-19 policy tables, sharing lessons learned through membership in the Variants of Concern Strategy Table, appearances before the House of Commons Standing Committee on Science and Research, and engagement with the Pan-Canadian Health Data Strategy Expert Panel.
- CanCOGeN’s mission-driven approach has been highlighted as an example of how to build a path forward for transformative innovation policy to tackle today’s grand challenges, as part of the Brookfield Institute’s [Canada’s Moonshot](#) report.

*“[The] Canadian COVID-19 Genomics Network, can serve as [a] model for the future by connecting public health leaders, university scientists, and governments across the country.”*

[Chief Public Health Officer’s Report on the State of Public Health in Canada 2021](#), p 83.



2021-02 Terrance Snutch, Karen Jones load nanosequencing device

## The future of genomic data sharing

As the COVID-19 pandemic has highlighted, there are significant needs for data sharing (genomic and metadata), a solid infrastructure and trained work force able to advance genomic surveillance across the country to address health challenges in Canada. This includes developing new ways to ensure equitable health benefits and data sharing for genomics and building systems to monitor and manage for future pathogen outbreaks.

- As noted in the recently released report [What We Heard: Pan-Canadian Genomics Strategy](#), data management and sharing is critical to leveraging genomics within Canada. Common data standards, privacy regulations, cross-jurisdictional coordination, and developing storage capacity and analytical capabilities are needed to enable broad access to and use of data across the genomics ecosystem. CanCOGeN laid the foundations for all of these components to build a strong future for pan-Canadian genomics data sharing.
- Directly building on our work with CanCOGeN, Genome Canada launched a strategic initiative on [Genomic Monitoring of Pathogens in Water](#) (GeMPaW).

While CanCOGeN focused on individual genomic surveillance, building capacity and sharing data, this initiative will focus on community genomic surveillance data to inform public policy. GeMPaW will scale up water-based genomic surveillance of emerging and antimicrobial resistant pathogens—driving adoption of best practices for wastewater testing, data sharing and pan-Canadian collaboration to inform policy and public decision making. It will also include specific funding for genomic surveillance projects in northern and remote communities that are led by community-based Indigenous researchers and/or organizations, take place on Indigenous land, and incorporate Indigenous knowledge.

- Finally, as noted above, as a direct result of CanCOGeN investments, there is ongoing repurposing of the new and existing sequencing capacity across the country to address other key public health challenges. For example, this infrastructure is now being used to sequence tuberculosis (TB), human immunodeficiency virus (HIV) and emerging pathogens in several provinces.



Biobanque au Saguenay, Lac Saint Jean, Génome Québec. Échantillons Photo

For more information, contact Koko Bate Agborosangaya, Director, Strategic Programs, Genome Canada, at [kagborsangaya@genomecanada.ca](mailto:kagborsangaya@genomecanada.ca)