



OPINION

Large-scale national data approach key to unlocking power of genomics

By NAVEED AZIZ MAY 4, 2022

Data-driven genomics research promises long-term benefits for Canadians and our health system: enhanced disease prevention, better predictions of future illness, and more accurate and personalized treatment options.



Minister of Innovation, Science, and Industry François-Philippe Champagne speaks to reporters following cabinet swearing-in on Oct. 26, 2021. A large-scale data-driven infrastructure approach at the federal level will offer returns far beyond those provided by study-specific government-funded research, writes Naveed Aziz. *The Hill Times* photograph by Andrew Meade

Genomic data is increasingly supporting new research across many sectors. To continue to maintain its scientific excellence in the new era of large-scale genomics, Canada needs a plan to translate the power of “big” genomic data into positive impacts for Canadians and the Canadian economy.

Genomic data is the information coded in DNA—the chemical sequence of As, Cs, Ts, and Gs that make up all life on earth. The last decade has seen the creation of data resources that merge and “bank” DNA sequences and individual-level data that represent a population (of any species—human, animal, plant, etc.) and yield the power to enable new discoveries through data-driven research. When it comes to genomic data, the refrain “data is the new oil” is fitting, but with one key difference: while oil is a diminishing resource, data has the ability to multiply and expand with use.



Dr. Naveed Aziz serves as chief executive officer of CGEn. Photograph courtesy of CGEn

Data-driven genomics research promises long-term benefits for Canadians and our health system: enhanced disease prevention, better predictions of future illness, and more accurate and personalized treatment options. While there is little argument about these impacts, the frameworks surrounding the use of big data are complex. The ability to create large enough data resources hinges on the capacity to generate representative genomic data at scale, as well as strategies to harmonize, collate, and share this data, along with demographic, lifestyle, and clinical information across geographic and institutional boundaries. Addressing these challenges, and capitalizing on the strong partnerships that exist across the health research landscape, enables a dynamic and thriving genomic data ecosystem that will lead us into a new age of genomic-based discovery.

Federal and provincial governments, research institutions, and other stakeholders are supporting this vision through investments into national infrastructures. For example, since its inception in 2017 as a Major Science Initiative, CGEn—Canada’s national platform for genome sequencing and analysis—has received support from the Canada Foundation for Innovation and other partners to enable and provide capacity for large-scale genomic sequencing data generation. Canada also boasts world-class genomics expertise. Canadian scientists play leadership roles in a number of successful disease-focused genomic research projects including The Cancer Genome Atlas (TCGA), International Cancer Genome Consortium (ICGC), Terry Fox Marathon of Hope Cancer Centres Network, and MSSNG Autism Project.

However, historically, Canada has lacked a national framework to access participant samples, manage data storage and stewardship, and navigate the ethics surrounding the sharing of sensitive human genomic data. In addition, Canada has yet to develop a plan for a truly national genomic

resource (akin to the United Kingdom's 100,000 Genomes Project or the All of Us Research Program in the U.S.) that fully represents Canada's rich diversity, and will ensure the delivery of science-based health care decision-making in the long term.

While the COVID-19 pandemic has been a devastating experience, it has also provided important opportunities. For one, the power and potential of genomics research surrounding COVID-19 is now evident to the general public and policy-makers alike. We have a chance now to build broad momentum and support for genomics research. Secondly, the need to quickly understand individual responses to COVID-19, in terms of disease severity, spurred a collaborative effort from institutions, funders, diverse clinical studies, and participants from across the country, enabling CGen and its partners to deliver the federally funded \$20-million COVID-19 Host Genome Sequencing project, [HostSeq](#). The resulting HostSeq databank will be Canada's largest national genomic databank to date, containing genome sequences from 10,000 individuals affected by COVID-19, linked with medical and clinical data that is easily accessible to investigators.

And while we were sure not to cut corners, we sure did cut some red tape—the HostSeq project ramped-up quickly, broke down previously insurmountable obstacles for interprovincial data governance, and found effective solutions to support Canadian science with Canadian data. As a result, numerous [research projects](#) are already benefiting from controlled access to the HostSeq databank. In addition, all HostSeq participants consented to broad sharing and use of their data, meaning the databank can continue to fuel other health research, beyond COVID-19. In short, HostSeq has created a national infrastructure as well as a resource that will support genomics research for years to come.

When it comes to a federally-supported national data approach, a clear strategy as well as standardized processes for genomic data generation, sharing and stewardship are needed. HostSeq has created the required infrastructure and lays the foundation for large-scale genomics efforts in Canada that will offer returns far beyond those provided by study-specific research. As the federal government makes decisions on the best way to implement the \$400-million Pan-Canadian Genomics Strategy announced in Budget 2021, it will be important to consider how Canada builds on this critical infrastructure that was previously lacking in Canada. Using HostSeq's success, we can, and should, maximize the strategic investment made by the federal government and significantly de-risk future efforts aimed at building synergistic national genomic resources that support and advance a variety of disease specific research areas.

CGEn's working philosophy is to operate in a ready state that can support projects at scale with fast turnaround times. The time is ripe for all partners and stakeholders to come together and build on the years of investment and effort towards a Canadian ecosystem where research is enabled by—and researchers have the ability to use—large genomic datasets as a resource.

The HostSeq blueprint, along with Canada's many strengths and the realization of the federal government's Pan-Canadian Genomics Strategy could help us capitalize on large-scale opportunities. That includes a Canadian population health genomic databank that will continually feed quality research programs and foster the development of talent and tools that are needed to succeed in the knowledge economy.

Dr. Naveed Aziz serves as chief executive officer of CGEn, Canada's national platform for genome sequencing and analysis, with nodes at The Hospital for Sick Children (Toronto), McGill University (Montréal) and Canada's Michael Smith Genome Sciences Centre (Vancouver).

The Hill Times

UPDATE: This article was updated on May 6, 2022.