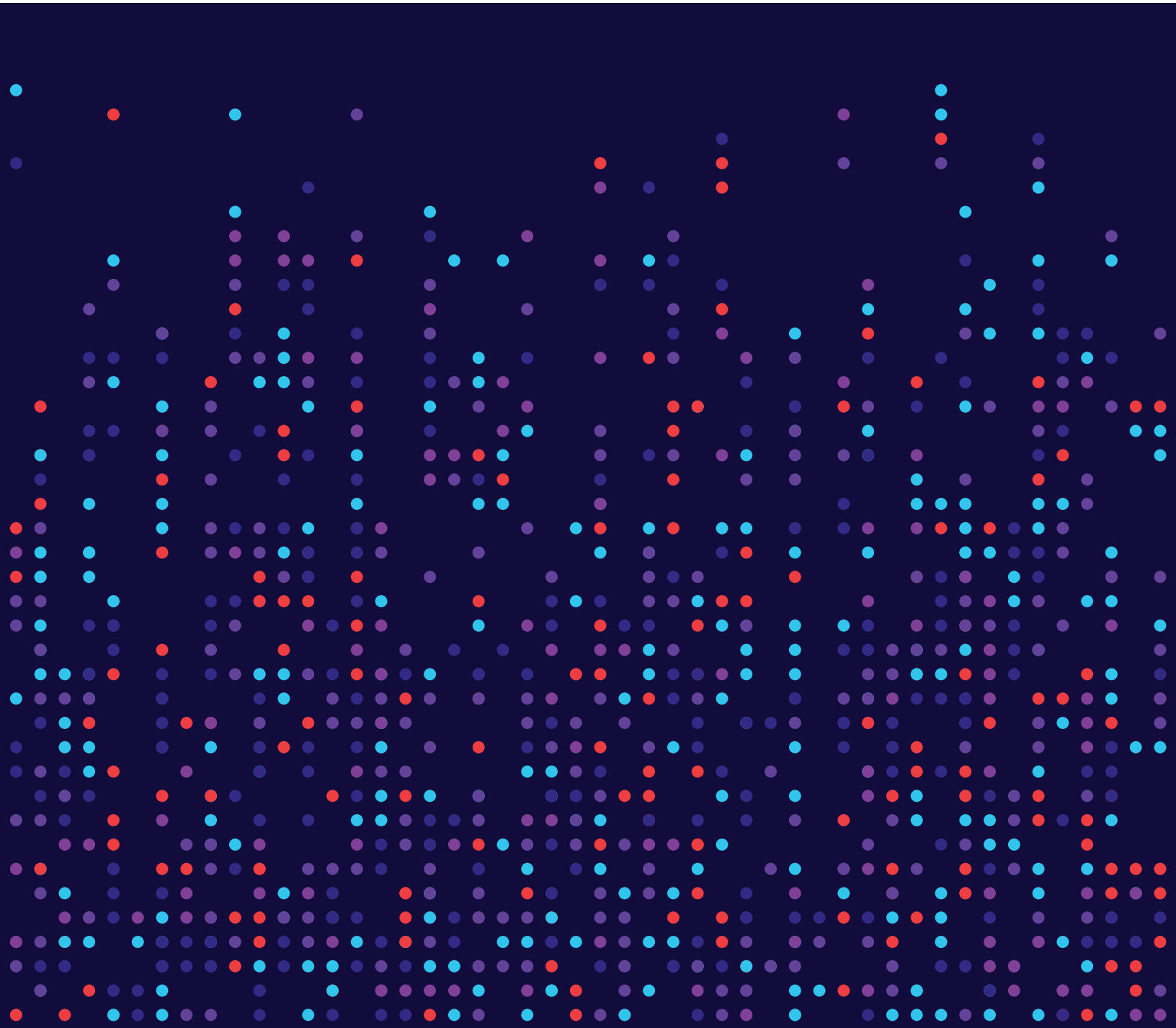




Impact Report



**The HostSeq Databank is supporting research
in many facets of COVID-19 and other
diseases – identifying biomarkers for disease
severity, predicting treatment response,
identifying genetic risk and protective factors,
and developing novel tools and techniques.**





Led by CGEn, HostSeq was established in 2020 with the federal government's support, as part of Genome Canada's Canadian COVID-19 Genomics Network (CanCOGeN). HostSeq brought together 15 COVID-19 research studies to sequence the genomes and gather clinical information for more than 10,000 people in Canada to understand host genetic contributions to COVID-19, and other health issues.

- 15 COVID-19 studies from across Canada
- 10,000+ participants

15 studies

Diverse COVID-19 contributing studies from 10 Canadian institutions

10,000+ participants

Participants from across Canada

50 staff

Staff engaged across 3 CGEn nodes

>1000 terabytes

Data produced across 3 CGEn nodes

18 publications

Research publications enabled by HostSeq (to date)

Canada's first accessible genomic health resource

HostSeq consent enables future data enrichment through linkage and participant re-contact

A blueprint for a Canadian large-scale genomics program

CGEn has demonstrated the capacity and capability, and developed the data governance, consent and study enrollment frameworks, to deliver population-level genomics initiatives

IMPACTS

Precision Medicine

- Using Canadian data for the discovery of new genetic markers to inform personalized treatments and therapies

Preparing for the Next Pandemic

- A data infrastructure that can be re-used for identifying genetic markers influencing susceptibility, severity, and outcomes

International Positioning

- Enabling Canadian contributions to global genomics initiatives

HostSeq Databank

10,000+ Records



Broad Consent

A Diverse Canadian Dataset



23 COVID-19 research projects using HostSeq data (to date)



150+ researchers and trainees analyzing HostSeq data (to date)



Inaugural cohort for the Pan-Canadian Human Genome Library



13 research projects in other health areas using HostSeq data (to date)



International collaboration through COVID-19 Host Genetics Initiative (HGI)

Summary

Led by CGEn, HostSeq was established in 2020 with the Canadian government's support, as part of Genome Canada's Canadian COVID-19 Genomics Network (CanCOGeN). HostSeq brought together 15 COVID-19 studies to sequence the genomes and gather clinical information for more than 10,000 people living in Canada, to fuel research into host genetic contributions to COVID-19 and other health issues. The resulting broadly-consented and diverse HostSeq dataset is enabling important national and international research, and through data enrichment, expansion and partnerships, the databank is positioned to be an important Canadian genomics resource for years to come. The delivery of HostSeq has also led to the development of data governance, consent and enrollment frameworks that constitute a blueprint for future Canadian population-scale genomic programs, critical for precision medicine applications, pandemic preparedness and Canada's international positioning.

HostSeq Implementation

HostSeq involved the active participation of 15 diverse contributing studies across 10 Canadian institutions, over 10,000 participants, and over 50 staff from CGEn's three nodes in Toronto (The Centre for Applied Genomics, The Hospital for Sick Children), Montreal (McGill Genome Centre, McGill University) and, Vancouver (Canada's Michael Smith Genome Sciences Centre, Provincial Health Services Authority). The implementation of HostSeq (Yoo et al., BMC Gen 2023) has led to the generation of over 1,000 terabytes of data and supported in various ways the publication of 18 papers.

Enabling Research in Canada and Internationally

The HostSeq Databank contains 10,000+ whole genome sequences (WGS) and clinical metadata records from a diversity of people living in Canada, that are consented broadly for sharing and research use.

Controlled-access to the HostSeq Databank is coordinated by the Data Access Compliance Office at the Centre of Genomics and Policy at McGill University, enabling investigators from the public and private sectors to solve critical health research questions using large-scale Canadian genomic data. A total of 36 projects have accessed HostSeq data since it became available in late 2021, supporting research activities of over 150 investigators and trainees. Twenty-three of these projects focused on COVID-19 research – identifying biomarkers for disease severity, potential treatment responses, genetic risk or protective factors, and more. There are 13 additional projects in other areas of health research including cancer, population genomics, autism spectrum disorder and cardiac health. Through its innovative design, HostSeq facilitated the return of WGS data to contributing studies, delivering over 10,000 sequences to support COVID-19 clinical research.

On the international stage, HostSeq was involved with the COVID-19 Host Genetics Initiative (HGI) – a worldwide effort to share and analyze data to uncover the genetic determinants of COVID-19 susceptibility, severity, and outcomes. The HGI identified 51 genetic locations significantly associated with COVID-19, mapped to three major biological pathways (The COVID-19 Host Genetics Initiative, Nature 2023). As HostSeq represented Canada's contribution to this important global body of research, this paper includes many Canadian researchers involved in HostSeq's delivery. Notably, a more recent analysis of the HostSeq dataset replicated two genetic loci reported to be associated with COVID-19 severity through the HGI, and identified associations with two additional genes to be further investigated (Garg et al., PLoS Genet. 2024).

To further support the use of HostSeq data and contribute to the training of genomic data scientists, CGEn actively sought partnerships in the Canadian ecosystem. Through the *CGEn HostSeq/Canadian Institutes of Health Research Joint Postdoctoral Fellowship Program* and the *CANSSI Ontario STAGE HostSeq Fellowships Program*, 9 fellows were supported in their COVID-19 research using HostSeq data, feeding a talent pipeline critical for advancing genomic medicine applications in Canada.

As the only existing Canadian genomic health dataset that is broadly consented for sharing, storage and use, HostSeq also represents the inaugural cohort for the CIHR-supported Pan-Canadian Genome Library (PCGL). The aim of the PCGL is to develop a responsibly governed infrastructure that unites human genome sequencing efforts across Canada, and acts as a key resource for the integration of genomics as a core component of learning healthcare systems.

Broad Consent

To enable broad use of HostSeq data for COVID-19 as well as other health studies, the HostSeq Databank has participant consent for:

- use by public and private sector researchers
- re-use for health research
- participant re-contact
- linkage to other health and administrative databases
- international sharing and use

Canada's First Accessible Genomic Health Resource

The HostSeq Databank was created to fuel COVID-19 and other health research broadly, in academic, industry and government sectors, using genomic data at a scale never-before available in Canada. CGEn has delivered on this goal and importantly, HostSeq can be seen as a perpetual data infrastructure. Future enrichment of the metadata, for example through linkage to administrative and other health datasets, and building on HostSeq's existing outputs through the PCGL will expand the breadth and depth of its research impact.

A Blueprint for Canadian Large-scale Genomics Programs

As a Major Science Initiative (MSI) of the Canada Foundation for Innovation (CFI), CGEn has a mandate to remain in a ready-state to respond to Canadian challenges. Through the successful delivery of HostSeq, CGEn has demonstrated not only its technical capacity, but also its capability in terms of expertise and operational management across the country, to support large-scale, complex genomics programs. Importantly, HostSeq's implementation helped to develop numerous foundational frameworks and processes that together constitute a workable blueprint for national genomic health databanking in support of Canadian research.

01

Data Governance

Development of this framework during a global pandemic provided an opportunity to cautiously but diligently push through some of the usual red tape to enable timely large-scale data generation, as well as access to this data. As a result of this work, the HostSeq DACO has prepared a policy document detailing development of the data governance framework and how the governance model was adopted to help translate research into practice. The policy document includes points to consider for the development of governance models and policy for population-level genomics initiatives both generally and in the event of future pandemics.

02

Minimal Consent Elements

A set of standardized HostSeq minimal consent elements was developed and utilized to enrol participants across all clinical studies, enabling broad data sharing, storage and use. The acceptance of these consent elements by more than 10,000 participants in Canada is an important demonstration to other cohorts, studies and biobanks interested in generating genomic data that broad consent can and should be used to enable future data re-use. As such, the HostSeq minimal consent elements were adapted into published guidance on consent in Canadian genomic studies (Longstaff et al., CMAJ 2022). This guidance, in turn, will be used to select cohorts and samples for inclusion in the Pan-Canadian Human Genome Library, in which CGEn is an applicant partner.

03

National Study Enrollment

HostSeq's study design enabled the enrollment of prospective and retrospective studies from across Canada that provided samples for sequencing at one of CGEn's three nodes. CGEn's close relationships with the research community supported the project's efficient implementation and HostSeq covered all sequencing costs. CGEn's effective delivery of HostSeq as a "study of studies" will facilitate the participation of contributing studies in future large-scale genomic programs. This design also benefits funders (CIHR and others) by enabling redirection of significant sequencing costs to other project activities.

Impacts

Precision Medicine

HostSeq data is already supporting research in many health areas, and the data remains available to Canadian and international researchers for future discoveries. HostSeq's 10,000+ genomes represent a rich, diverse pilot for the provision of large-scale genomic data in Canada that can inform essential and powerful precision medicine research. However, the current databank must be expanded beyond its current status and scale to maximize its utility. Many countries have built genomic programs to ensure they can advance precision medicine with research based on large-scale data (including the UK – 1 million WGS, US – 1 million WGS, Israel – 100,000 WGS, Denmark – 60,000 WGS). Utilizing the HostSeq blueprint, CGEn is committed to working with funders and partners to grow HostSeq to an internationally competitive scale.

Pandemic Preparedness

The manifestation of disease for any future pandemic pathogens are expected, like COVID-19, to be dependent on an individual's unique genetic makeup. As new infectious diseases emerge, it will be necessary to act quickly in order to understand the role of human genetics, deliver new biomarkers for risk prediction, and implement appropriate healthcare strategies including potential precision therapies. HostSeq's existing data infrastructure can be enriched with new pandemic-specific data through participant recontact or supplemental data linkage. In leveraging the investments made through HostSeq, Canada will assure it is response-ready to enable timely, critical research for future pandemics. HostSeq data is already supporting the identification of markers that impact COVID-19 susceptibility, severity and outcomes, which may also play a role in other infectious diseases.

International Positioning

The generation of large-scale genomic data through HostSeq has enabled Canada to play a part in the COVID-19 Host Genetics Initiative (HGI)'s worldwide efforts to characterize COVID-19 and prepare for future pandemics. The inclusion of HostSeq data in the HGI demonstrated the willingness of Canadian participants, institutions and governments to share individual-level data outside of the country. This collaboration with international initiatives positions Canada as a key player in population studies at a global scale and demonstrates that Canada, with CGEn's genome sequencing capacity and capability, is willing, primed and ready to generate and share high-quality data resources.



HostSeq was a part of Genome Canada's Canadian COVID-19 Genomics Network (CanCOGeN), funded through the Federal Government.



GenomeCanada

