10 YEARS OF **IMPACT**

CGEn

CGEn

Canada's national platform for genome sequencing and analysis

A Major Science Initiative of the
Canada Foundation for Innovation

Land Acknowledgement

CGEn spans across the traditional and unceded territories of many nations including the Musqueam, Squamish, and Tsleil-Waututh Nations, the Huron-Wendat and Petun First Nations, the Seneca, and the Mississaugas of the Credit River, and on land which has long served as a site of meeting and exchange amongst Indigenous peoples, including the Haudenosaunee and Anishinabeg nations. These sites are home today to many diverse First Nations, Inuit, and Métis people, and CGEn is grateful to be able to share this land.

It is important for us to take pause, acknowledge and reflect on the impacts of colonialism across Indigenous communities in Canada, particularly in how colonization has and continues to influence genomics and genetic research. We remain eager to hear stories, humble to learn lessons, and ready to continue fostering meaningful and inclusive change.

Reflecting on CGEn's 10 years of impact

In celebration of a decade of leadership in genomic science, ***CGEn: 10 Years of Impact*** showcases how CGEn's national collaboration continues to drive transformative research across Canada and beyond. Founded in 2015 by Canada's three leading genomics centres, CGEn established one of the world's largest genome sequencing and analysis infrastructures, expanding the scale and ambition of Canadian research and placing our country at the forefront of genomic science.

Building on the strengths of our nodes and our country's collective genomics expertise, CGEn has evolved into a nationally coordinated, regionally distributed platform for research, with strong international connections that reinforce Canada's role in global initiatives. With support from the Canada Foundation for Innovation, host institutions, provincial governments, Genome Canada and other partners, CGEn has continued to advance alongside rapidly evolving genomic technologies. By implementing and optimizing new tools and methods, developing innovative services, and engaging highly trained staff, CGEn has ensured that Canadian research remains at the leading edge by generating high-quality and sovereign genomic data for discovery.

CGEn's infrastructure and leadership have supported and advanced major national initiatives such as HostSeq, the Canadian BioGenome Project, and the Terry Fox Marathon of Hope Cancer Centres Network, while serving more than 3,000 academic, industrial, and government laboratories across sectors from health and agriculture to conservation.

These efforts have generated thousands of publications, accessible genomic datasets, and invaluable training experiences that will fuel the future impact of genomics in science, policy, and innovation.

Guided by CGEn's Vision 2030 **'Every biological study powered by genomic data—driving a healthier, more sustainable future for all Canadians'**, CGEn remains committed to ensuring researchers have access to world-class technology, data, and expertise. This progress is made possible through the commitment of our staff, leadership, funders, and partners across Canada's research ecosystem. We thank the CGEn Board of Directors, Scientific Advisory Board, and all collaborators for their continued partnership in advancing genomic science. Together, we celebrate a decade of impact and a future of discovery driven by shared purpose and genomics innovation.

Respectfully,

Gordon C McCauley
Chair, CGEn Board of Directors
President & CEO,
adMare BioInnovations

Meredith McLaren
CGEn CEO

Contents

Our DNA: Mandate & Milestones

Return to Contents

About CGEn

CGEn is Canada's national platform for genome sequencing and analysis. Established in 2015, CGEn employs over 200 staff and is funded primarily by the Canada Foundation for Innovation (CFI) through its Major Science Initiatives Fund (MSIF) and Innovation Fund (IF), leveraging investments from the provincial governments of Ontario, Quebec, and British Columbia, Genome Canada, host institutions, and other sources of funding. CGEn operates as an integrated national platform with nodes in Toronto (The Centre for Applied Genomics at The Hospital for Sick Children), Montréal, (McGill Genome Centre at McGill University) and Vancouver (Canada's Michael Smith Genome Sciences Centre at the Provincial Health Services Authority). CGEn leads large-scale projects and provides advanced genomic services to enable research by over a thousand Canadian and international research groups every year, spanning the public and private sectors, including those working in human health, agricultural sciences, fisheries and oceans, ecology, biodiversity, and many other disciplines.

CGEn Impact at a Glance

3 nodes and **220+ staff** serving

550+ institutions **3,000+ laboratories** **1,000s of research projects**

Advancing Canada's most impactful genomics research programs through leadership, collaboration and high-quality data

Canadian Biogenome Project · HostSeq · Marathon of Hope Cancer Centres Network · CARE forRARE · Canadian Precision Health Initiative

15.8 petabases
of data generated

3.4 petabases
publicly accessible for re-use

3,500+ publications
by staff and users

250+ international
co-authorships with CGEn scientific leaders

Driving research, development, application & commercialization across many important sectors

CGEn Vancouver · CGEn Montreal · CGEn Toronto

Technology Development for emerging technologies, enhancing service and developing new tools

Training the next generation of genomics leaders to keep pace with large-scale genomics and its growing applications

Our Partners & Funders

Innovation Canada Foundation for Innovation · Provincial Health Services Authority · SickKids · McGill · Genome Canada · British Columbia · Ontario · Québec

Vision

Every biological study powered by genomic data—driving a healthier, more sustainable future for all Canadians.

Mission

Strengthen Canada's national capacity for genome sequencing and analysis, enabling impactful regional, national, and international research with large-scale, high-quality data.

Accelerate Canadian genomic science and solidify Canada's global reputation by uniting the country's top genomics expertise.

Anchor strategic collaborations across initiatives, funders, companies, and organizations, positioning Canadian genomics initiatives for the greatest possible impact.

Priority Areas

1. High-Quality Data At-Scale
2. Technology Development
3. Talent & Skills Development
4. Genomics Community Engagement

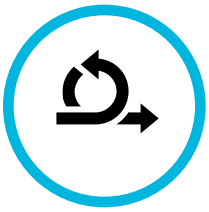
Values

Excellence

We are committed to the highest standards of scientific rigor and service, ensuring the maximum quality in everything we do.

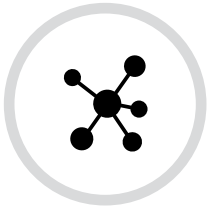
Connectivity

We foster value-driven connections within our network, and with the broader scientific community to advance research and its societal benefits.



Agility

We remain adaptable and responsive to emerging scientific, technical, and societal challenges and opportunities, ensuring timely and relevant action.

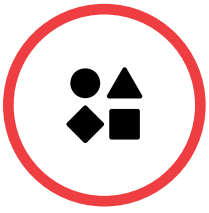


Reach

We leverage our national platform and partnerships to extend the impact of Canadian genomics and advance collaborative initiatives that drive innovation and benefit society.

Innovation

We push the boundaries of genomic science, integrating novel ideas and technologies to drive practical solutions and breakthrough discoveries.



Diversity

We support inclusive research practices and foster a diverse and enriching work environment. We enable genomics research across diverse sectors, organization types, and regions.

Highlights of the CGEn story so far

CGEn nodes share a history of innovation and pioneering of genomics technologies. Through ongoing collaboration and information sharing within our platform and wider network, CGEn draws on the unique expertise of each node to support operations and consolidates leading scientific perspectives to benefit the broader Canadian research community. CGEn's distributed structure is also ideal for supporting national initiatives where institutional, funder, or other requirements require local sequencing.

Our nodes: A heritage of impactful genomic science

1998 The Centre for Applied Genomics (TCAG) established

TCAG was founded in 1998 as a core facility at The Hospital for Sick Children (SickKids), supporting in its early years the Human Genome Project and other genome research locally, across Canada, and worldwide. Under the leadership of its co-founder and Scientific Director, Dr. Stephen Scherer, TCAG has supported a wide variety of research with an emphasis on the genetic and genomic basis of human variability, health and disease, including research on the genetics of autism spectrum disorder (ASD) and related neurodevelopmental conditions, and structural variation of the human genome. This research contributed to the identification of genome-wide copy number variations (CNVs), including defining CNV as a highly abundant form of human genetic variation[1]. In 2007, Scherer and colleagues launched the Personal Genome Project Canada, a resource that supported evaluation of whole genome sequencing as a tool for personalized medicine[2]. In 2008, TCAG became the first Canadian centre (and one of the first in the world) to operate all three major "next generation" DNA sequencing (NGS) technologies of the time (Illumina Solexa, Applied Biosystems/Life Technologies and Roche/454). Scherer's advocacy with the Canadian Coalition for Genetic Fairness helped to establish Canada's Genetic Non-Discrimination Act, which passed into law on May 4, 2017.

1999 Canada's Michael Smith Genome Sciences Centre (GSC) established

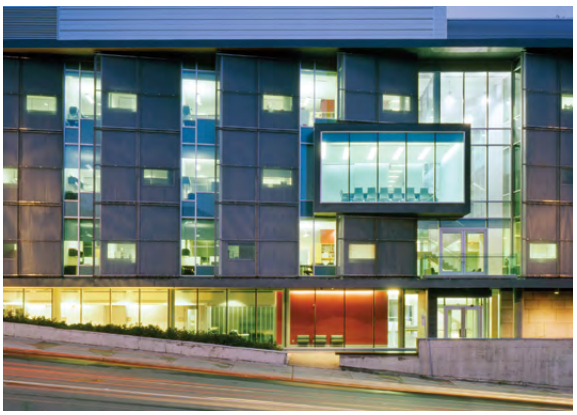
CANADA'S MICHAEL SMITH GENOME SCIENCES CENTRE

The GSC opened its doors in 1999 as the world's first sequencing centre embedded in a cancer clinic (BC Cancer) and has since applied genome technology and bioinformatics expertise to populations of cancer patients, identifying risks for hereditary cancers and helping to prevent, diagnose and treat many different cancers in entirely new ways. Under the leadership of Dr. Steven Jones and Dr. Marco Marra, the GSC has been a pioneer in many national and international genomics efforts. GSC was an early participant in the Human Genome Project, has built reference genomes for a diversity of species, was the first in the world to sequence the genome of the SARS coronavirus (2003)[3], and received one of the world's first NGS machines in 2006 (the Solexa 1G/Illumina Genome Analyzer). GSC represented Canada in The Cancer Genome Atlas (TCGA) consortium as one of six Genome Characterization Centers, characterizing more than 10,000 tumour samples using genomics and providing RNA-specific data analysis support for the initiative. GSC scientific contributions include the first cancer sequenced to inform clinical decision making[4] and the first breast cancer genome[5].

2002 McGill Genome Centre (MGC) established

McGILL GENOME CENTRE

The MGC opened in 2002 to provide Canadian and international researchers with high-throughput technologies and cutting-edge approaches to enable next-generation genomic studies. MGC scientists helped to initiate and complete the International HapMap (Haplotype Map)—a resource that describes the common patterns of sequence variation across the human genome—and pioneered the application of these results for large-scale investigations to identify the molecular contributions to human disease. Founder and inaugural Scientific Director Dr. Thomas Hudson pioneered the introduction of NGS to Montreal, and MGC scientists have since made ground-breaking advances in the use of genomics for the study of monogenetic disease and other disorders. Under the current leadership of Dr. Mark Lathrop, MGC has empowered study designs as well as sequencing and analytical solutions for genome variation, provided comprehensive tools to assess functional differences in human tissues and cells, and developed scalable techniques and robust interrogations of functional genomic variation in populations. MGC is a partner in the landmark program DNA to RNA: An Inclusive Canadian Approach to Genomic-based RNA Therapeutics (D2R), aimed at creating an international hub for a global research effort specializing in the development and delivery of more inclusive genomic-based RNA therapeutics and is part of the Victor Phillip Dahdaleh Institute of Genomic Medicine at McGill, established in 2023. The centre is leading important research on the ethical, policy, and legal implications of genomic medicine and works closely with provincial and federal agencies on issues of health surveillance and monitoring.



Major CGEn Milestones

2015 **CGEn is formed.** The Canada Foundation for Innovation (CFI) Innovation Fund (IF) brings the three major genomics centres together to form CGEn, Canada's national platform for genome sequencing and analysis ($58.4 million*)

2017 **CFI Major Science Initiatives Fund (MSIF) awarded** to support CGEn's national platform operations ($42.9 million)

CGEn's CanSeq150 launches in honour of Canada's 150th birthday, sequencing new genomes to support conservation and biodiversity research

2020 **CFI IF awarded** to renew and enhance CGEn infrastructure ($30 million*)

CGEn's HostSeq Initiative launches with partners across the country, in response to the COVID-19 pandemic ($20-million investment by Innovation, Science, Economic Development (ISED) Canada through Genome Canada's CanCOGeN Initiative)

HostSeq

2021 **Canadian BioGenome Project launches** with CGEn nodes providing collaborative technical support in the generation of high-quality reference genomes for important Canadian species (funded by Genome Canada and other partners)

CANADIAN BIOGENOME PROJECT

2023 **CFI MSIF renewed** to support CGEn operations ($48.9 million). Genome Canada Genomic Facilities Technology Development funding (linked to MSIF) awarded to CGEn ($3.3 million)

CFI IF awarded to further renew and enhance CGEn infrastructure ($19 million*)

2025 **Canadian Precision Health Initiative (CPHI) launched** by Genome Canada to build a genomic data resource containing 100,000 human genomes, with CGEn selected as the sequencing partner for most CPHI projects

Canadian Precision Health Initiative

* total funding from CFI, provincial governments and other partners

Select Achievements

First Canada beaver genome generated and made publicly available through CanSeq150

95 species sequenced across CGEn nodes, paving the way for the Canadian BioGenome Project

10,000+ high-quality whole human genomes sequenced

36 research projects supported by the HostSeq databank

Inaugural dataset for the Pan-Canadian Genome Library

340 species sequenced across CGEn nodes

New methods developed for challenging non-human samples

By 2025, CGEn has:

Processed 525,000+ samples

Sequenced 155,000+ whole genomes

Generated 15.8 petabases of data

Scaling Canadian Genomics

Return to Contents

Leading Genomics Technology Innovation in Canada

Early technology investments in CGEn made by the Canada Foundation for Innovation (CFI) enabled Canada to join other G7 countries in the ability to perform accurate, cost-effective genome sequencing and build a national platform critical for advancing Canadian genomic science. CGEn's scientific expertise, iterative CFI funding, and strategic investments from other funding partners, have enabled CGEn to succeed in being Canada's leader in genomic data service provision and technology development.

Consolidated expertise for Canadian genomic science

Expertise and collaboration across CGEn's national infrastructure enable Canadian researchers to conduct genome-wide experiments with high precision and reliability. Our scientific leadership who are amongst the top, deep technical knowledge and strong partnerships enable CGEn to generate and interpret high-quality data that drive discovery and innovation across Canada's genomics ecosystem.

Stephen Scherer hosting pioneering genomics researcher J. Craig Venter at CGEn-Toronto

220+
Expert CGEn staff

63%
Technical roles: hands-on service provision, operating & maintaining infrastructure

21%
Scientific roles: data handling, processing and preparation for analysis

10%
Management: oversight of laboratory and compute infrastructure, study design, facility operations

6%
Administrative support roles

Comprehensive genomics services for the research community

Genomics encompasses more than reading DNA bases – CGEn's comprehensive services provide high-quality data to inform investigations across the complexity of biology.

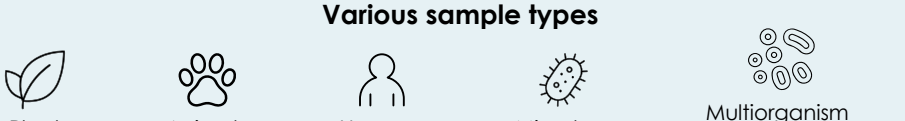
PROJECT CONSULTATION

CGEn's expert staff works with researchers to:

- Understand research questions
- Select sequencing and data analysis strategies
- Establish timelines

SAMPLE RECEIPT AND QUALITY CONTROL

Various sample types



DNA
encodes genetic instructions

Whole genome sequencing – reads an organism's entire genome, giving the most complete picture of the genetic code

Metagenome sequencing – reads the genomes of all organisms (typically microbial) in a sample

Exome and targeted sequencing – reads protein-coding regions or selected gene targets

Epigenome sequencing – reads chemical tags that control whether genes are turned on or off

RNA
helps read, carry and regulate genetic instructions

Transcriptome sequencing – detects which genes are active inside cells

Metatranscriptome sequencing – detects which genes are active inside cells

Spatial transcriptomics – uses sequencing to map which genes are active and their location within intact tissues

Protein
performs the functions encoded by genetics

Proteomics (sequencing-based) – identifies and measures proteins in a sample

DATA GENERATION USING CGEn INFRASTRUCTURE

CGEn has processed more than **525,000 samples**, including more than **155,000 for whole genome sequencing**

Single-cell approaches – apply sequencing at the level of individual cells identifying information that bulk cell analyses can miss

BIOINFORMATICS AND DATA RETURN

15.8 petabases of data analyzed and returned to users

Turning raw data into high-quality, meaningful information for Canadian Science

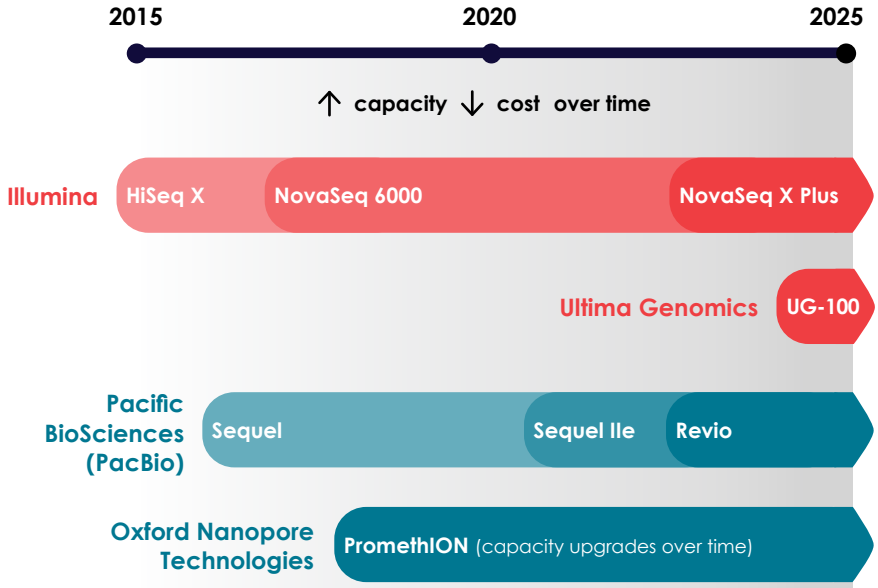
CGEn infrastructure: keeping pace with research needs

Continual testing and implementation of emerging technologies ensures that CGEn has the capacity to meet increasing demand for large-scale genomics, and that Canadian research benefits from the highest quality data at internationally competitive pricing.

CGEn's Major Sequencing Technologies

Short-read genome sequencers read small fragments of DNA, enabling relatively rapid, accurate and cost-effective understanding of the genome.

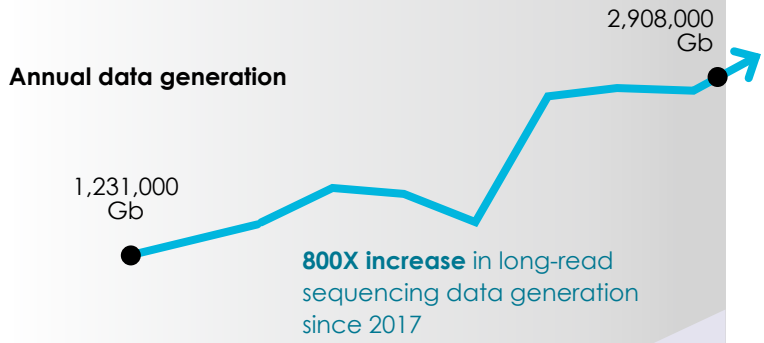
Long-read genome sequencers read much longer stretches of DNA, allowing better resolution of complex regions, structural changes and assembly of more complete high-quality genomes.



Data Generation from all CGEn services

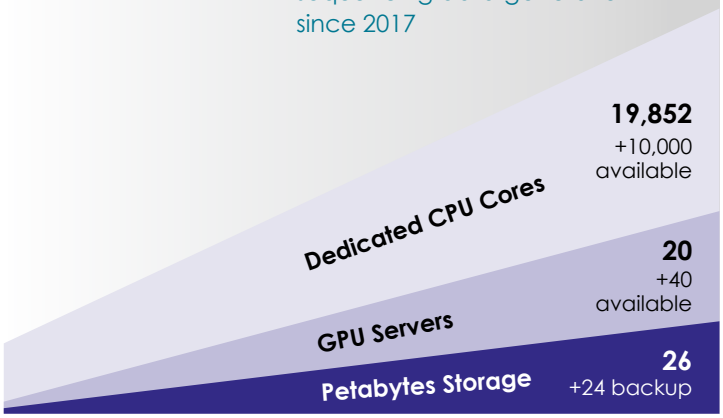
15,774,000 gigabases

Equivalent to 200,000 4k movies = 48 years of binge-watching!



CGEn Compute and Storage Capacity

Compute and storage capacity are required for data processing and management of increasingly large genomic datasets.



Technology Development: Optimizing infrastructure, processes and services

Benchmarking emerging technologies in preparation for national scale projects

Benchmarking is an essential CGEn activity that ensures quality, accuracy, reliability and consistency of data, and is critical for national projects where data from multiple nodes are reassembled into downstream datasets.

In 2020, CGEn nodes compared Illumina short-read whole genome sequencing using standard controls, resulting in at least 95% agreement in detecting and over 99% accuracy in correctly identifying genomic variants[6]. Successful benchmarking experiments like these show that genome sequencing projects can rely on the quality and reproducibility of aggregate data generated across CGEn's distributed nodes. The most significant differences in the data were introduced by the nodes' analysis pipelines, which led to updates across the sites to further increase standardization. This Illumina benchmarking was particularly important for CGEn's rapid response to research needs surrounding the COVID-19 pandemic, where CGEn created a dataset of more than 10,000 high-quality whole genome sequences as part of the HostSeq initiative. CGEn's Technical Expert Committee continues to benchmark data from emerging technologies being applied in large-scale studies, such as long-read sequencing, to ensure our platform remains in a ready-state to support Canadian genomic interests.

Pictured above: CGEn Technical Experts Group, Executive Committee and Scientific Advisory Board (Vancouver, April 2024)

Enhancing CGEn service and genomic sciences

Over the last 10 years, CGEn technology development activities have been critical in enabling more efficient and affordable sequencing services and developing tools that result in broader benefits to the genomics research community. Some examples are described below.

Uncovering cellular diversity through single-cell genomics innovation

Single-cell genomics supports investigations into development, disease and cellular functions at a level that bulk analyses cannot capture, and has been an important area of focus for CGEn's technology development activities. CGEn-Vancouver has developed novel methods and improvements to separate single cells from tissue samples, sequence whole transcriptomes, small RNAs, and obtain epigenomic signatures. These methods have been applied in CGEn service, enabling collaborators and other researchers to more comprehensively profile single cells in tumours and obtain deeper insight into tumour composition, disease progression, and other mechanisms that are key to improving cancer outcomes[7,8]. Results and protocols from this work were ultimately used to develop the Cancer Single Cell Dynamics Observatory (CFI-funded infrastructure led by S. Aparicio and housed at CGEn-Vancouver), which will develop new single-cell measurement approaches and ultimately provide cancer patients with improved diagnostics and therapeutic options. CGEn-Montreal has focused on developing protocols for preparation and sequencing of single nuclei[9,10], spatial transcriptomic methodologies[11] and the application of cost-effective sequencing technology in large-scale single-cell atlases[12], work that is continuing by exploring the new Ultima Genomics platform.

Improving efficiency through automated sample preparation

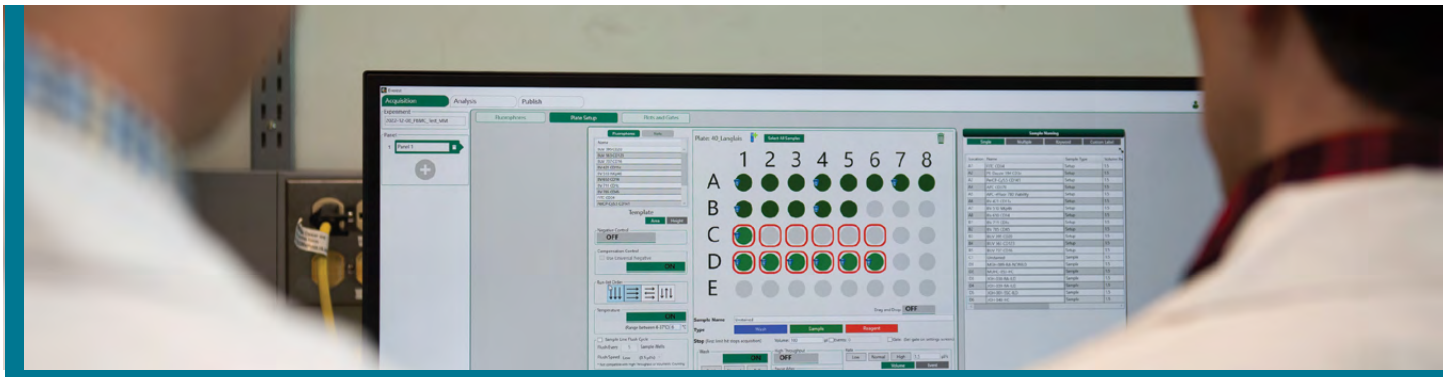
CGEn continues to lead in automated sample preparation, largely due to in-house, integrated instrumentation and robotics expertise. At CGEn-Vancouver these improvements have resulted in the ability to flexibly sequence different batch sizes without impacting sequencing efficiency in long-read sequencing (which enables high-throughput projects), and the development of automatable methods for isolating cell-free tumour DNA from blood. At CGEn-Montreal, work with INTEGRA Bio (and formerly Miroculus) has resulted in microfluidic sample preparation protocols for long-read sequencing, and a new generation instrument is being tested to enable fully automated high-throughput long-read sample preparation. These efforts have improved consistency of results, reduced operator fatigue and injury, decreased costs, increased efficiency and allowed CGEn to better contribute to large-scale genomics projects.

Enhancing reliability with innovative sample tracking controls

CGEn-Vancouver has developed a sample tracking method using unique controls created by inserting random sequences into a common DNA backbone[13]. By adding these controls to tissue or DNA samples upon receipt, they are processed and sequenced alongside each sample. At the data analysis stage, sample identity can be confirmed and cross-contamination detected—alleviating the necessity for standard QC testing.

Building species' complete genomes for the first time

CGEn has led optimization methods for *de novo* genome assembly (or producing a complete genome without a reference – like building a puzzle without the picture on the box) through large-scale projects CanSeq150 and the Canadian BioGenome project, and other related work. At CGEn-Toronto iterative approaches have led to increasingly complete and accurate mammalian genome assemblies, starting with initial work on the Canada beaver[14], progressing to the Wolverine[15] and Muskox[16]. Development of these methods using non-human species is intended to ultimately lead to routine, low-cost, high-throughput applications with human biomedical utility. CGEn-Montreal has integrated high-quality genome assembly generation with transcriptomic data for improved annotation and gene discovery, enabling the development of effective insect pest control strategies[17,18,19,20].



Bioinformatic tool development

CGEn has developed several bioinformatic tools, used in-house for specific projects and made available to researchers across the globe to benefit genomic science broadly.

CGEn nodes have a strong historic track record in developing widely used bioinformatic tools and methods: the CIRCOS viewer that provides a graphical synopsis of genomic changes and alterations (original paper[21] cited more than 10,000 times), the ABySS assembler which was the first of its kind capable of producing mammalian-sized whole genomes, and GenPipes that facilitates multi-step genomic workflows. More recent examples include:

- LocusFocus22, a suite of software tools specializing in data integration. LocusFocus helps researchers understand why a specific genetic region (or "locus") found in genome studies might be important, by testing whether the locus also impacts nearby genes being switched on or off. This provides information about the biological mechanisms behind potential genetic-disease associations. LocusFocus has been adopted by many researchers worldwide—in 2024-2025, it recorded 985 unique visits.
- A comprehensive workflow for identification of copy number variation[23] (which involves duplications or deletions of DNA segments) using WGS data. The workflow employs 6 algorithms and was used to detect rare mutations in individuals with autism spectrum disorder.
- NanoMethPhase[24], a tool used in Parent of Origin Genome Aware and Rapidomics projects, that uses long DNA sequencing reads to determine which parts of the genome inherited from each parent are methylated (chemically modified). This provides information about gene activity differences between the maternal and paternal copies of DNA across large regions of the genome.
- vikNGS[25], a software package to analyze sequencing data and indicate whether genetic variants are linked to traits or diseases, even when the data is noisy or uneven in quality.
- MAVIS[26], a structural variant caller used by the Personalised OncoGenomics, PROFYLE, Marathon of Hope Cancer Centres Network projects and for users broadly.

Fuelling Innovative Science

Return to Contents

Genomics-powered research is improving lives and shaping a sustainable future for Canada.

As Canada's national platform for genome sequencing and analysis, CGEn leverages its expertise and infrastructure to lead major Canadian initiatives and contribute to national and global projects, strengthening Canada's international standing in genomics. CGEn also serves thousands of researchers across academia, industry, government, and non-profit sectors. Through leadership, collaboration and service, CGEn is helping to advance the biological understanding of human disease, population health, biodiversity, environmental resilience and more.

From fundamental discovery to applied innovation, CGEn's activities continue to deliver benefits across science, society, and the economy.

CGEn's national userbase

CGEn supports researchers across scientific disciplines, organization type and geographic location.

550+ unique organizations
3,000+ laboratories

Location

Canadian
91%

International
(spanning 36 countries)
9%

Affiliation

Academic
85%

Government or not-for-profit
8%

Industry
7%

Primary focus

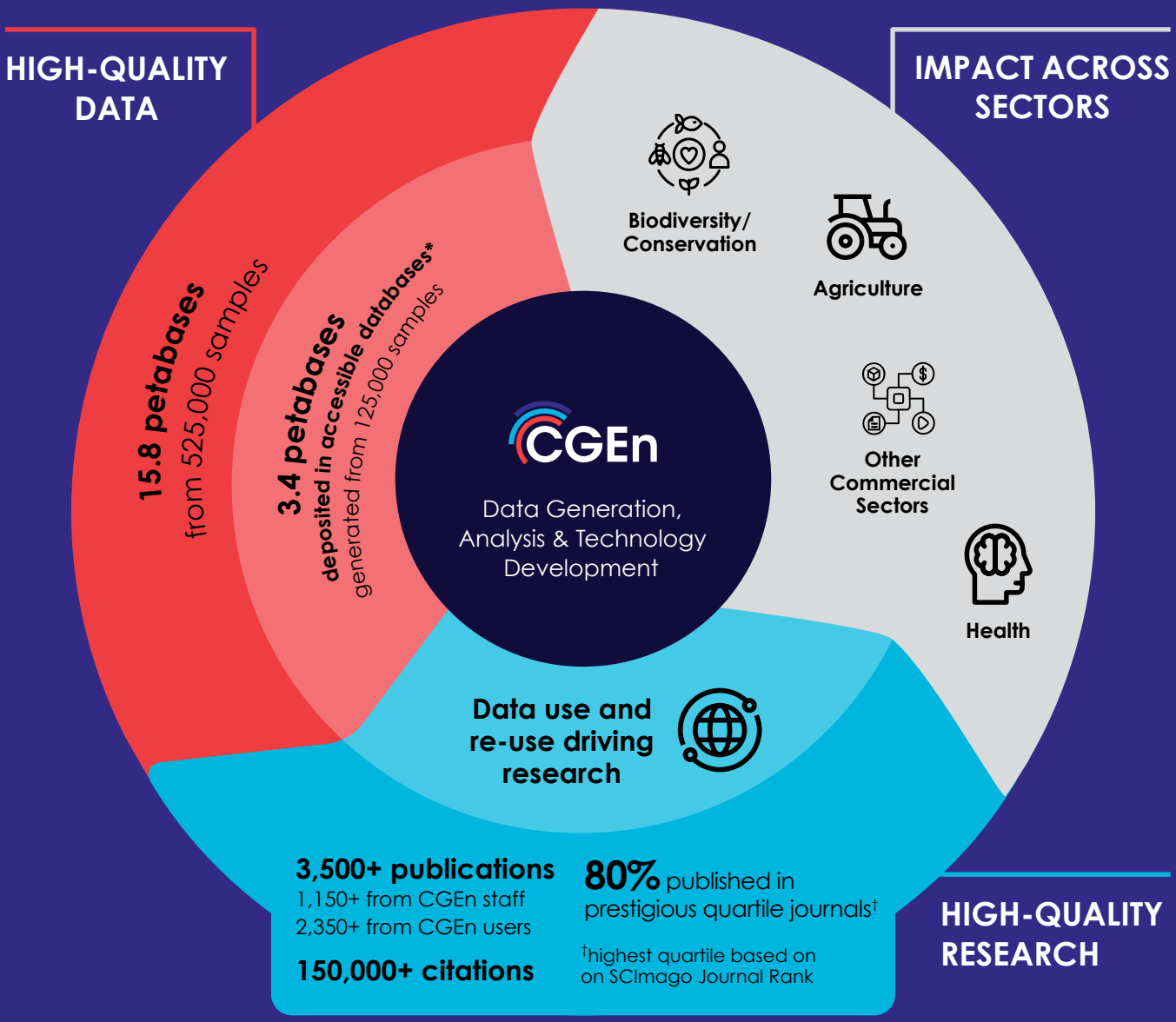
Health and Human Studies
75%

Biodiversity and Conservation
13%

Commercial Sectors
12%

CGEn: From data to impact in Canada and beyond

CGEn-generated data and technology development have supported thousands of publications and other outputs that inform a wide range of research and applications. Beyond this, CGEn champions making genomic data accessible for future research when ethical and legal requirements are met. Broadly-accessible data increases return on research investment and amplifies impacts on patients, the environment, biodiversity, and the economy.



* Data are deposited on: ARGO, Brain-CODE, Calcul Québec, cgHub, NCBI dbGAP, DNA Data Bank of Japan, Epigenomics Data Analysis and Coordination Center, European Genome-Phenome Archive and Canadian Genome Archive, European Nucleotide Archive, NCI Genomic Data Commons, GenBank GEO, HostSeq, MSSNG, NCBI Short Read Archive.

Ecosystem Impact

Canadian solutions for health data management

Genomics health research involves a complex array of research ethics and consent, study design, experimentation, and disseminating massive and often highly sensitive datasets to multiple users, databases, and partners. CGEn's leadership and collaborators are thought leaders in these areas and have been major contributors to Canadian data management efforts.

Our staff and infrastructure support internal database design and implementation, genetic variant calling and analysis, data transfer and data storage, with integrated security and privacy best practices, helping to build a Canadian genomics ecosystem that responsibly meets the needs of the research community now and in the future.

The Canadian Genome-Phenome Archive (CGA)

The Canadian Genome-Phenome Archive (CGA) was developed over several years and established in 2025 at CGEn-Vancouver. The CGA provides a national service for the permanent archiving and distribution of genomic, demographic and clinical data generated by Canadian health research projects. Where participant consent is established, the CGA archives and enables release of research data, under strict protocols aligned with best practices of the Federated European Genome-Phenome Archive (fEGA), of which it is the first non-European node. The CGA will enable visibility and use of Canadian genomic datasets globally, marking a major milestone for our country's international position in biomedical research and personalized medicine. The CGA is also connected to the Pan-Canadian Genome Library (PCGL), which sets out a national data management system for human genomic data. The PCGL is hosted at McGill University, linked to CGEn's Montreal node.

The Database of Genomic Variants (DGV)

The Database of Genomic Variants (DGV) works with the US-based National Centre for Biotechnology Information and the European Variation Archive to actively curate and disseminate copy number and structural variation data from healthy individuals. Established in 2004 by The Centre for Applied Genomics (now CGEn-Toronto), the DGV continues to be hosted at CGEn-Toronto where it is continuously updated with new studies. As of August 2025, it contains about 7 million variations, representing 988,000 locations in the genome, derived from 75 studies that include about 47,000 individuals. The DGV has more than 15,000 unique visitors on average each month and helps researchers understand whether a given structural variant is typical in the human population, supporting genomic research and facilitating hundreds of thousands of clinical interpretations each year.

Canadian Distributed Infrastructure for Genomics (CanDIG)

Canadian Distributed Infrastructure for Genomics (CanDIG) is a fully distributed platform allowing researchers to run genomics analyses on a national scale, while keeping data private and under local control. CanDIG's leadership includes G. Bourque (associated with CGEn-Montreal) and CGEn-Vancouver Scientific Director S. Jones. CGEn-Montreal and CGEn-Vancouver host two of the three CanDIG instances in Canada, helping to build and maintain the platform as well as curate and submit data. CanDIG participates in Common Infrastructure for National Cohorts in Europe, Canada, and Africa (CINECA), leading efforts of the Global Alliance for Genomics and Health (GA4GH) Discovery Networks and Researcher Identity streams that ensure interoperability with European colleagues.

The Northern Biobank Initiative (NBI)

The Northern Biobank Initiative (NBI) has the potential to empower northern and rural populations across British Columbia with world-class research opportunities. Acknowledging the importance of sequencing and data sharing, CGEn-Vancouver leadership lent expertise to the setup and ethics application for the initiative's first prospective biobank (Northern BC Biobank), which was approved in 2025. CGEn-Vancouver will receive samples from the NBI for sequencing and provide secure data storage for this endeavour.

Dr. Nadine Caron (University of British Columbia and Northern Medical Program), NBI project lead, speaks at the NBI Phase 2 launch.

CGEn expertise for international collaboration

CGEn is driving international connections through its provision of high-quality genomics services, support of global initiatives and the expertise of its scientific leadership. CGEn's national genomics infrastructure ensures that Canada can support domestic projects and build capacity, while enabling international participation to expand knowledge, coordination and opportunity for impact.

280+ International user laboratories from 36 countries

250+ International co-authorships with CGEn scientific leaders

Some global initiatives advanced through CGEn service, data and expertise:

HostSeq

COVID-19 hg

CANADIAN BIOGENOME PROJECT

EARTH BIOGENOME PROJECT

Canadian Genome-phenome Archive

Federated European Genome-phenome Archive

RNA Cooperative

THE MEDICAL GENOME INITIATIVE

RNome

PGC Psychiatric Genomics Consortium

TCGA

IPCHiP International Precision Child Health Partnership

TARGET 2035 LIGAND-AI

Select Exemplar Research Projects

enabled by CGEn

Health Sector

Return to Contents

Health Sector

Genomics is transforming our understanding of disease, with research providing the basis for effective and personalized interventions. From discovery science to clinical translation, CGEn has enabled human and health research through its scientific, technical and data governance expertise. Over 80% of whole genome sequences generated by CGEn have been human, 75% of CGEn users have health as their primary area of focus, and our scientific directors and associated investigators are leaders in the genetics of cancer, autism, and many other disease areas.

Large-scale population cohorts combining genomic and participant data, and **focused studies** in specific disease areas, are driving advances in screening, treatment, and prevention, to power precision medicine. By supporting studies that link genetic changes—both inherited and acquired—to disease, CGEn is playing a key role in advancing human health research in areas such as **cancer, rare diseases, neurological and cardiovascular disorders, and many more.**

Over **80%** of whole genome sequences generated by CGEn are **human**

75% of CGEn users have **health** as their primary area of focus

Select Health Research Areas

- POPULATION COHORTS
- CANCER
- NEUROLOGICAL CONDITIONS
- RARE DISEASES
- ADDITIONAL HEALTH AREAS

CGEn's HostSeq Databank

The COVID-19 pandemic provided an opportunity to demonstrate Canada's capacity and capability to implement a population-scale genomics project. In response to this public health emergency CGEn conceptualized and implemented the Host Sequencing Initiative "HostSeq", in collaboration with 15 Canadian clinical studies and other partners across the country, to understand how variations in the human genome contribute to COVID-19 and other conditions.

HostSeq generated a databank of over 10,000 whole genome sequences with matched demographic and clinical information broadly accessible for research studies with approved ethical oversight[27]. Supported by Innovation, Science and Economic Development Canada through Genome Canada's Canadian COVID-19 Genomics Network (CanCOGeN), HostSeq implementation was led by a multidisciplinary committee of investigators including CGEn Scientific Directors S. Scherer, M. Lathrop, and S. Jones, L. Strug (associated with CGEn-Toronto), N. Aziz (former CGEn CEO), B. Knoppers (associated with CGEn-Montreal), and S. Turvey (University of British Columbia) and supported by numerous specialized working groups. **CGEn and these committees were responsible for oversight of clinical study enrolment, genomic data generation at one of CGEn's three nodes, large-scale genomic data analysis, clinical data standardization, and establishment of data governance and sharing principles in collaboration with McGill University's Centre of Genomics and Policy.**

HostSeq was initialized within months of the declaration of the COVID-19 pandemic in 2020, and more than 10,000 short-read genomes (representing ~525,000 Gb) were sequenced by March 2023. CGEn compiled and analyzed the resulting data on an ongoing basis over this timeframe, ensuring approved researchers had timely access to data as it became available.

HostSeq used a broad consent strategy that enables public and private-sector researchers to access data for COVID-19 and any future health research, data linkage and international data sharing. Through HostSeq, CGEn helped develop numerous foundational frameworks and processes that together constitute a workable blueprint for national genomic health databanking and a pilot for the provision of large-scale genomic data in Canada that can inform essential and powerful precision medicine research.

CGEn's HostSeq Databank

Research impacts of HostSeq

36 approved projects

150 investigators and trainees

25 projects identifying biomarkers for disease severity, potential treatment responses, genetic risk or protective factors related to **COVID-19**

11 projects focused on **other health research** including cancer, population genomics, autism spectrum disorder and cardiac health.

Examples of research projects using HostSeq data:

- An analysis of the HostSeq dataset replicated two genetic loci previously associated with COVID-19 severity through the international COVID-19 Host Genetics Initiative and identified two new genes to be further investigated[28].
- Findings from a project using HostSeq data revealed associations between specific Human Leukocyte Antigen (HLA) variants and COVID-19 disease severity, which may inform responses to other infectious diseases[29].
- S. Scherer's team used the HostSeq dataset as a population control in conjunction with autism spectrum disorder (ASD) cohorts (Autism Speaks MSSNG, Simons Simplex Cohort, and Simons Foundation Powering Autism Research) to analyze 400,000+ X-chromosome variants, identifying over 50 associated markers and 14 genes with significant ASD-associated variants, eight of which are novel associations[30].

CGEn has a mandate to remain in a ready-state to respond to Canadian challenges. As new infectious diseases emerge, it will be necessary to act quickly to understand the role of human genetics, deliver new biomarkers for risk prediction, and implement appropriate healthcare strategies including rapid screening and potential precision therapies. Leveraging the investments made through HostSeq, for example enriching the dataset with new pandemic-specific data, Canada can enable timely, critical research. In addition, collaborations with international initiatives like the HGI position our country as a key player in population studies at a global scale and **demonstrate that Canada, with CGEn's genome sequencing capacity and capability, is willing, primed, and ready to generate and share high-quality data resources.**

CGEn's HostSeq Databank: Leading and delivering a complex national genomics project

Ecosystem Impact

Through its innovative "study of studies" design, HostSeq funded sequencing and facilitated the return of genomic data to contributing studies to support COVID-19 clinical research, supporting a number of publications addressing study design and genetic variation associated with COVID-19[31,32,33,34].

HostSeq data was a Canadian contribution to the international COVID-19 Host Genetics Initiative, which identified 51 genetic loci significantly associated with COVID-19 susceptibility, severity and outcomes[35].

The acceptance of HostSeq's broad consent elements demonstrated that the Canadian public generally supports secure storage and ethical sharing of genomic data for future research. As such, the HostSeq minimal consent elements were adapted into published guidance on consent in Canadian genomic studies[36].

CGEn's joint fellowship programs with the Canadian Statistical Sciences Institute and the Canadian Institutes of Health Research enabled nine post-doctoral fellows and graduate students to apply cutting-edge computational and statistical genetics methodology to the HostSeq data. As Canada moves toward a learning health system based on genomic and related health data, training on large-scale data is critical for feeding the talent pipeline that will advance genomic medicine applications.

HostSeq represents an inaugural cohort for the Pan-Canadian Genome Library (PCGL; funded through the Canadian Institutes of Health Research). HostSeq is helping develop PCGL data ingestion, analysis, and access processes. This work will also inform the deposit of more than 90,000 WGS from CGEn to the PCGL through the Genome Canada-funded Canadian Precision Health Initiative (CPHI; running from 2025 to 2029), and other future initiatives.

Silent Genomes Project

The Silent Genomes Project is addressing systemic and structural barriers that limit equitable access to genomic health care for Indigenous populations in Canada.

Led by L. Arbour, W. Wasserman and N. Caron (University of British Columbia), the project's main goals included the identification and breaking down of barriers to precision diagnosis for Indigenous patients, the establishment of governance structures guided by Indigenous priorities, and the development of an Indigenous Background Variant Library (IBVL). **The IBVL was developed using short-read WGS data generated at CGEn-Vancouver from 596 consented individuals across four communities** participating in the First Nations arm of the Canadian Alliance for Healthy Hearts and Minds study. It supports more accurate clinical interpretation of genetic variants for Indigenous people by providing information on the frequencies of specific genetic variants in generally healthy First Nations participants which can be compared to genetic test results. Prior to sample transfer, Indigenous governance structures were established to guide all aspects of the work, including frequency of specific variant release, manuscript review, and research application review. These processes were formalized in alignment with Indigenous Data Sovereignty principles.

The IBVL was launched online on January 31, 2025, and there are now over 90 approved users, primarily physicians, genetic counsellors, and laboratory scientists, utilizing the IBVL to aid in clinical diagnosis of Indigenous patients. An example of a clear clinical benefit of this work is the DPYD Screening webpage, which provides genomic variant frequency data gleaned from the IBVL to improve outcomes for Indigenous patients undergoing cancer treatment.

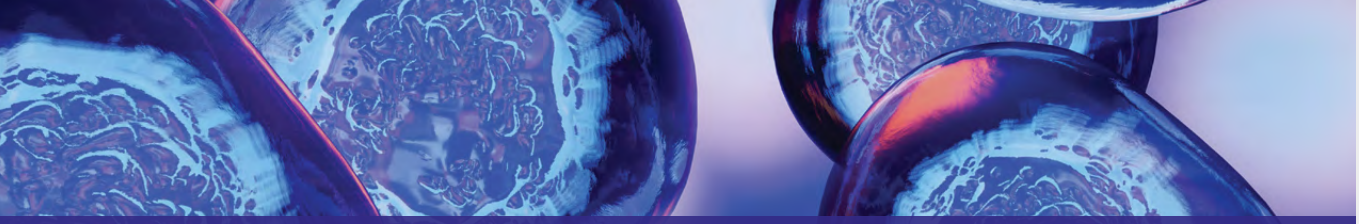
Canadian Longitudinal Study of Aging (CLSA)

Supported by CIHR, CFI, and other partners, the Canadian Longitudinal Study of Aging (CLSA) is following 50,000 individuals between the ages of 45 and 85 across Canada, for at least 20 years. CLSA is led by P. Raina (McMaster University), with C. Wolfson (McGill University) and S. Kirkland (Dalhousie University).

Since 2011, CLSA has collected information on changing biological, medical, psychological, social, lifestyle and economic aspects of peoples' lives and made this data available for research. These factors are being studied to understand how they impact both health maintenance and the development of disease and disability as people age. **CGEn-Montreal processed samples from over 23,000 CLSA participants, and performed whole-genome genotyping to identify variations** in specific DNA locations, then used a reference dataset to fill in missing information[37]**.** These genotyping data were linked to CLSA's comprehensive compendium of data collected over time, opening up new opportunities to integrate genetics into health and ageing research. The linked data was made available in 2018 through the CLSA data access application process. Over 750 research teams have requested CLSA data access thus far, and 73 projects have published research based on CLSA genotyping data. Analysis of the linked data, along with other international datasets, uncovered over 100 novel loci associated with key parameters to define glaucoma[38] and also validated the contribution of a genetic-based risk score to screen individuals with high fracture risk[39].

This valuable data resource can also support direct identification of common genetic variations associated with conditions related to complex traits. As an example, the large-scale genotyping data has been linked to psychosocial factors to investigate both independent and interactive effects on cardiovascular disease[40].

Social BEACON (Social disparities-Biological ExplorAtion CONnecting multi-omics approach with healthy aging), funded by CIHR and led by P. Raina, will introduce new measures into the CLSA, such as epigenetic analyses on additional CLSA participants, and proteomics, which measures proteins in the blood. CGEn-Montreal is now focusing on generating cost-effective, whole genome epigenetic studies through advanced long-read sequencing approaches in collaboration with ONT and carrying out a pilot study for detection of somatic mutations (acquired over a person's lifetime, for example in cancer cells) based on technology development with Ultima Genomics.



Personalized OncoGenomics (POG)

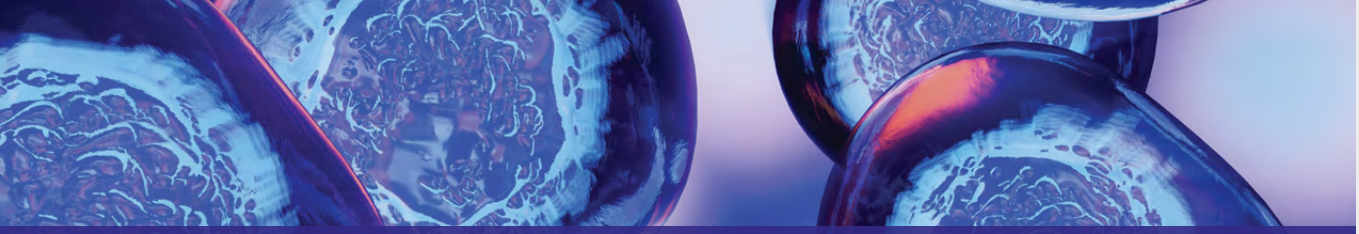
BC Cancer's Personalized OncoGenomics (POG) program is a world-leading research initiative that uses intensive genomic data generation (at CGEn-Vancouver), collection and analysis to identify and better understand the genomic alterations that drive cancer growth, metastasis, and response to therapies in poor prognosis cancer patients[41].

Led by J. Laskin and M. Marra (associated with CGEn-Vancouver), POG integrates genomic, transcriptomic, clinical, and treatment data, providing a rich dataset for research. As of August 2025, POG has enrolled over 2,276 participants and short-read WGS has been completed for 1,576 paired tumour and normal (non-tumour) participant samples. "Paired" sequencing is important in cancer research to identify pre-disposing hereditary mutations in normal tissues and "subtract" mutations from the tumour sequencing results that are also found in normal tissue to focus on tumour mutations. Almost all samples include transcriptome data, and 511 have associated long-read WGS. In addition, the project has returned more than 1,650 genomic results to patients and their oncologists to inform clinical care.

Since its inception, more than 60 papers associated with POG have been published, adding proof that genomics can fundamentally change the way cancer is treated.

Highlights of POG research:

- A study including 189 patient tumour POG samples showed that long-read sequencing may reveal additional clinically relevant information in patients whose tumours contain complex genetic alterations that cannot be decoded with short-read sequencing data. The authors have made all long-read data, and computational tools and algorithms used to analyze it, available to the research community to help stimulate further research in this field[42].
- A study analyzing the data of the first 570 POG patients revealed DNA evidence of drug resistance and cancer progression as a result of therapies, demonstrating the potential for precision oncology to provide more effective therapeutic options for advanced cancer patients[43].
- The discovery and characterization of gene fusions in some cancers helped to identify new, effective therapeutics in these cases. These discoveries are now being transitioned into clinical use to benefit patients. Software has also been developed at CGEn-Vancouver to accurately detect these and other genetic structural variants[44,45,46]).

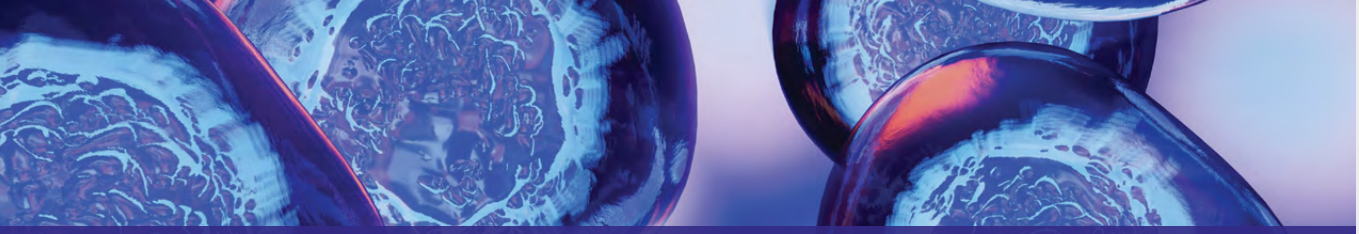


The Terry Fox Precision Oncology For Young People (PROFYLE)

PROFYLE is a project for children and youth with cancer who have not responded to standard treatment, giving them another chance through genomics.

The PROFYLE program is led by a national team of experts including researchers associated with CGEn nodes: D. Malkin (PROFYLE Program Director, SickKids), N. Jabado, I. Ragoussis and G. Bourque (McGill, associated with CGEn-Montreal), CGEn Scientific Director S. Jones and M. Marra (BC Cancer, associated with CGEn-Vancouver). The initiative represents a combined 1,500+ children with hard-to-treat cancers by uniting three Canadian paediatric cancer genomics legacy programs BC Pediatric Personalized Oncogenomics (Peds-POG), Ontario's SickKids Cancer Sequencing Program (KiCS), and Quebec's Personalized Targeted Therapy in Refractory or Relapsed Cancer in Childhood (TRICEPS) program. **With CGEn support (whole genome, exome and RNA sequencing), PROFYLE has demonstrated the utility of genomics in cancer detection and precision therapies for these patients.**

KiCS identified findings positively impacting outcome in 56% of cases. 54% of patients had at least one variant targetable by a specific treatment, and 6% received a modified diagnosis[47]. A Peds-POG study using transcriptome and tumour genome sequencing data generated by CGEn-Vancouver from 79 cancer patients showed that 12-15% of children with cancer have inherited mutations that predispose them to cancer, many of which are not identified in the clinic[48]. New funding through the Terry Fox Marathon of Hope Cancer Centres Network, and other programs, is expanding on these programs.



The Terry Fox Marathon of Hope Cancer Centres Network (MOHCCN)

MOHCCN is a national network launched in 2019 that includes more than 50 member institutions, 100 funding partners, and 800 individuals who are working together under a single vision: cancer patients in Canada will receive personalized cancer treatments tailored to the genetic makeup of their cancer – no matter where they live – leading to better outcomes and quality of life.

Building on principles and expertise developed through POG and PROFYLE, MOHCCN has reached its initial goal of sequencing 15,000 tumour/normal pairs to create a Gold Cohort, which will power precision oncology research with large amounts of high-quality, well-annotated data[49]. **CGEn nodes are the sequencing partners for the BC and Quebec consortium members and, as of February 2026, have generated and transferred data for 8,000 cases, including tumour and normal short-read WGS, to MOHCCN's Gold Cohort.** MOHCCN complements several existing Canadian programs—for example, a partnership with POG has expanded the study's enrolment to 7–10 patients weekly and POG is now leading MOHCCN Phase 2 initiatives for the return of participant results. Collaborative technology development work, funded by the Terry Fox Research Institute, is ongoing between CGEn-Montreal and CGEn-Vancouver associated scientists (G. Zogopoulos, J. Ragoussis, G. Bourque, M. Marra, K.Schrader) to explore the use of long-read technologies to identify as yet uncovered variants and link them to cancer predisposition, in preparation for integration of these exciting new technologies in future large-scale projects.

MOHCCN recently received $80M in new funding as part of the Canadian Budget 2025, which will increase the size of the Gold Cohort to 26,000 cases over the next four years.

Ecosystem Impact

Informing cancer genomic data sharing across Canada

In addition to providing services for the initiative, CGEn-Vancouver and CGEn-Montreal actively participate in many of the MOHCCN working groups, contributing data generation and data sharing expertise. MOHCCN has successfully deployed CanDIG, a distributed infrastructure for national genomic analyses, at three consortia sites, British Columbia (hosted at CGEn-Vancouver), Ontario and Quebec. CGEn-Vancouver was the first to upload genomic and clinical data to the CanDIG platform, supporting the development of technical and procedural frameworks needed for seamless data exchange within the network. This has helped MOHCCN to demonstrate the feasibility of interprovincial sharing of highly annotated cancer genomic data, and provide policies and guidelines that can inform genomic data sharing in the broader Canadian context.

Translating cancer genomics into greater clinical impact

At CGEn-Vancouver, population-level comprehensive characterization of tumours through large-scale projects, such as POG and MOHCCN, have generated a great deal of genomic data that is linked to clinical and other phenotypic data. To aid the translation of these insights into greater clinical impact, CGEn-Vancouver has developed multiple in-house databases to organize and house these data, including a variant database with curated publications and clinical trial data for genomic interpretation, and a clinical database, (CHAMP) that houses patient clinical information such as patient history, therapy and outcome data. Together, these databases will allow the controlled interrogation of vast genomic data generated by POG and other projects to inform future studies.

Diagnosing Ovarian and Endometrial cancers Early (DovEE)

DOvEEgene is a new genomic uterine pap test developed to screen for ovarian and endometrial cancer.

DOvEEgene was developed by a team of McGill researchers led by L. Gilbert. Initial work using samples from women undergoing pelvic surgery showed that the DOvEEgene test could detect 70% of cancers of the ovaries, fallopian tubes and endometrium with 100% accuracy. The test is highly sensitive—it can find one cell with mutations in its DNA within a sample of 1,000 normal cells. These differences are then analyzed by an AI algorithm which identifies whether the identified abnormal cells are cancerous, enabling detection of cancer in earlier stages. **CGEn-Montreal supported this work by sequencing 5,500 patient samples and assisting with data analyses and data management.** An ongoing clinical research study will now determine whether a genetic test of Pap samples taken from the uterus is capable of detecting ovarian and endometrial cancers early. Samples for this study will also be processed at CGEn-Montreal with the aim to translate the assay into a clinical test.

Parent of Origin Aware Genome Analysis (POAga) project

Around 300,000 Canadians are at high risk for developing hereditary cancers and may not know it.

The POAga project is led by I. Schrader (associated with CGEn-Vancouver), CGEn-Vancouver Scientific Director S. Jones, and P. Lansdorp (BC Cancer), in collaboration with the Hereditary Cancer Program of BC Cancer. POAga uses a cutting-edge technique based on long-read sequencing and strand-seq (method that reads one DNA strand at a time to help map chromosome organization) to pinpoint the parental origin of genetic mutations driving cancer. Using only a blood sample from a cancer patient, this project will validate POAga in conditions such as hereditary breast and ovarian cancer and Lynch syndrome, rarer syndromes with parent-of-origin-effects, and other genes predisposing patients to breast and gastrointestinal cancers. So far, **288 samples have been sequenced using nanopore long-read and other cutting-edge sequencing technologies at CGEn-Vancouver, with a planned total of 685**. A new spin-off startup, Evident Genomics, was launched last year based on patented POAga work conducted at CGEn-Vancouver.

Identifying the familial source (father or mother) of a patient's inherited risk for disease can improve follow-up screening of other family members, thereby increasing the chances of preventing or catching cancer early on.

Autism Genome Project (MSSNG)

MSSNG was launched in early 2016 with the goal to provide the best resources to enable the identification of many subtypes of autism, which may lead to better diagnostics, as well as personalized and more accurate treatments.

MSSNG continues to build on many advances in autism spectrum disorder (ASD) genomics, gene discovery, clinical diagnostics, and gene list development[50,51,52,53]. It constitutes one of the world's largest collections of genomic data for the study of ASD, containing short-read WGS and detailed phenotypic information for over 13,800 individuals from families affected by ASD. **Supported by data generation at CGEn-Toronto and led by CGEn-Toronto Scientific Director S. Scherer, MSSNG is a collaboration between Autism Speaks, Verily, DNAstack, and clinicians across Canada, North America and internationally (including the Qatar autism genome sequencing study).** It leverages the Simons Foundation Autism Research Initiative data (New York, USA) for comparative and validation studies and collaborates with Autism Innovative Medicine Studies-2-Trials (Europe) WGS efforts. MSSNG also contributes to the Government of Canada's Digital Supercluster, aiming to create the first federated, global network for sharing genomics and clinical data to accelerate discoveries and the development of precision therapeutics.

MSSNG is used by 370 research scientists from 72 institutions and companies in 20 countries around the world.

In 2022, Scherer's team published a comprehensive description of the genomic architecture in ASD using the MSSNG database. In this study, 135 ASD-associated genes of clinical diagnostic value – 68 entirely new – were identified[54]. In a recent study from Scherer's group, clinical testing data along with MSSNG and other autism data was utilized to identify rare damaging variants in the gene DDX53 in multiple families with autism, providing evidence this gene is linked to ASD and should be considered in clinical genetic testing[55]. Additionally, data from MSSNG was used to highlight that 50% of families who received WGS results through research reports had at least one clinically relevant finding related to ASD that led to a genetic diagnosis and counselling benefits[56].

In October 2025, Scherer published an opinion piece, 'The truth of autism will continue to be found in science, not myth' in *The Globe and Mail*, challenging U.S. administration claims about autism and stressing the need for evidence-based research over misinformation.

Genome Sequencing in Cerebral Palsy

Understanding the genetics of cerebral palsy (CP) can offer hope for improved diagnosis and treatment through precision medicine approaches.

In a seven-year study, representing a collaboration between groups in Ontario, Quebec, and Alberta, CGEn-Toronto generated short-read WGS data for 327 children and their family members. This study illustrated that more than one in ten children with CP had a genetic variant associated with their condition, suggesting that CP results from a combination of genetic and environmental factors[57].

The study's open data initiative, hosted on the Ontario Brain Institute's Brain-CODE analytics platform and the European Genome-phenome Archive, provides a valuable resource for future work by other researchers, aiming to identify new genes and pathways involved in CP and ultimately transform treatment options for affected children and their families.

Neuro Genomics Partnership

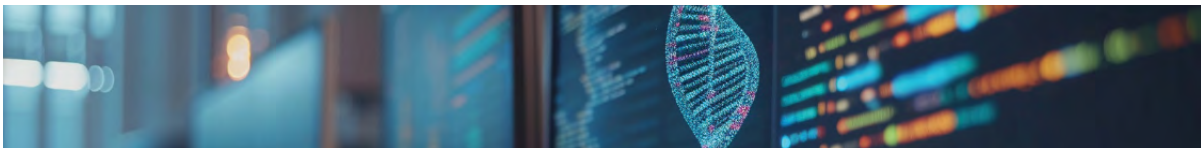
The Neuro Genomics Partnership is focused on understanding the genomics of patients with Parkinson's Disease and REM sleep behaviour disorder (RBD).

The Neuro Genomics Partnership is a collaboration between The Neuro (Montreal Neurological Institute-Hospital), Takeda Development Center Americas, Inc. and Roche. For this initiative, **CGEn-Montreal has generated WGS data for more than 2,200 patients which is currently being analyzed.** The ultimate goal of the study is to use genomic sequencing to characterize the conversion of RBD to Parkinson's Disease or Dementia and select candidates with faster conversion for clinical trials for these disorders. Over the past decade, CGEn-Montreal has also completed targeted sequencing for Parkinson's Disease and other disorders in over 10,000 patients through the Quebec Parkinson Network and the Canadian Open Parkinson Network, as well as other cohorts worldwide, including from the Mayo Clinic, which has led to the validation of 100 genes known or suspected to be involved in Parkinson's.

AI and Genomics

Artificial intelligence is transforming genomics from a data-rich science into a discovery engine. CGEn integrates machine learning into data analysis workflows to better detect and make sense of genetic differences, co-develops advanced tools with partners and invests in high-performance GPU computing to power these data-intensive models at scale.

When used in research alongside vast genomic datasets, AI can accelerate the discovery of new biological targets, clarify what genes do, and help researchers generate new lines of inquiry. CGEn produces genomic data at the scale and quality level required to turn biological complexity into actionable and accurate insights.



Examples of AI-enabled research supported by CGEn:

Using POG and other datasets, researchers developed and validated Supervised Cancer Origin Prediction Using Expression (SCOPE), a set of novel machine-learning techniques that use RNA sequence information to identify primary tumours[58]. The method had approximately 99% accuracy in identifying primary cancers and a success rate of 80-86% in the most challenging cases that had already failed human assessment or were extremely difficult to diagnose by a human expert.

Whole genome sequencing data identified rare genetic variants in children with cerebral palsy, which were then used with AI-based protein modelling tools to examine how these changes might alter protein interactions or disrupt key cellular signalling processes[57]. This analysis shows how AI can help move from identifying DNA changes to understanding their potential biological impact.

Samples sequenced through the KiCS program were used to validate a machine learning algorithm as the basis for a tool that can quickly and precisely identify specific cancer subtypes and increase accurate diagnoses in the clinic[59].

Panel discussion "Open questions and opportunities for AI in genomics" moderated by CGEn-Toronto Associate Director Lisa Strug at the collaborative SickKids/University of Toronto AI in Genomics Symposium, August 2025

Care4Rare Canada Consortium

As approximately 80% of rare diseases (RDs) have a genetic basis, analyzing an individual's genome can help pinpoint the underlying cause, leading to faster and more accurate diagnoses. This is especially important given that the 'diagnostic odyssey' for rare diseases often spans years and involves multiple specialists and misdiagnoses.

Over the past decade, CGEn has been supporting the Care4Rare Canada Consortium, made up of 200 physicians and 100 scientists focused on improving the diagnostic care of RD patients in Canada and around the world. Based out of the Children's Hospital of Eastern Ontario (CHEO) Research Institute, Care4Rare includes 21 academic sites across Canada and is led by K. Boycott (CHEO), M. Brudno (University Health Network), J. Majewski (McGill University; associated with CGEn-Montreal), L. Armstrong (University of British Columbia), and C. Marshall (SickKids; associated with CGEn-Toronto). The systematic identification of genes involved in RD has been a major success of the Care4Rare Consortium through overlapping projects including the flagship Finding of Rare Disease Genes in Canada (FORGE, funded by Genome Canada in 2011). FORGE identified over 65 causative (or suspected) genes and resulted in over 30 publications. The project received renewed funding through March 2017 for the Enhanced CARE for RARE Genetic Diseases in Canada (Care4Rare, subsequently C4R) and again through March 2024 as Care4Rare-SOLVE. Each of these studies was designed to address the challenges of the previous phase, and participating families remain in the Care4Rare program until their RD is solved.

Care4Rare Consortium leverages exome, genome and other 'omic data generation services at CGEn-Toronto to drive the identification of the molecular cause of unsolved RDs, achieving a diagnostic yield of 34% (623/1,806 of participating families), including the discovery of deleterious variants in 121 genes not previously associated with disease. The Consortium continues to study candidate variants in novel genes for 145 families[60].

In 2025, the consortium launched its latest initiative, C4R-EXPAND. As part of Genome Canada's Canadian Precision Health Initiative (CPHI), this pan-Canadian project will generate and share 17,650 diverse rare disease datasets—including short-read genome, long-read genome, transcriptome, and methylome data— aiming to provide new genetic diagnoses for thousands of families. CGEn-Toronto is playing an active role in this latest initiative as the project's main sequencing partner.

RapidOmics – Genetic Disorders in Infants

Genetic disorders are a leading cause of major illness and death in infancy. Long-read genome sequencing offers the promise of greater diagnostic sensitivity and shorter turnaround time for urgent diagnosis of genetic disease, which can profoundly influence clinical management and have lifelong consequences for parents and families.

Led by J. Friedman (UBC) and C. Ivany (BC Provincial Health Services Authority), **RapidOmics is using rapid long-read sequencing at CGEn-Vancouver to test 100 BC patients, either acutely ill infants or pregnant women with a fetus at very high risk of genetic disease.** The project is assessing the clinical value, limitations, costs and benefits of this type of sequencing as a diagnostic test. To date, 30 trios (parents and affected child) have been processed at CGEn-Vancouver and analyses of the long-read sequencing datasets were returned within a clinically useful timeframe. This research will ultimately improve the quality of health care provided to babies with genetic diseases and their families in BC, with the hope to one day extend this service to more than 3,000 pregnancies or infants annually across Canada.

Genomics for Cardiovascular Disease

Cardiomyopathy is an inherited heart condition that impacts up to one in 500 individuals. The condition affects the structure and function of the heart and can ultimately lead to heart failure.

SickKids researchers S. Mital and R. Yuen led the first-ever study to use WGS to examine tandem repeat expansions (TREs)—a form of genetic structural variation where short DNA sequences are repeated many times—in heart conditions. **Instead of focusing only on well-studied genes, they used short-read whole-genome sequencing, targeted long-read sequencing and advanced bioinformatics (developed in collaboration with CGEn-Toronto) to scan the entire genome for repeat expansions,** paying particular attention to enhancer regions (parts of the genome that act like switches, turning nearby genes on or off). The team found that people with cardiomyopathy are more likely to carry large or unusual tandem repeats in enhancer regions compared with healthy individuals. Enhancers control gene activity, so these repeat expansions may change the way heart-related genes are expressed, providing a new explanation for why cardiomyopathy develops in some people and potentially leading to better diagnostic tools and targeted therapies[61].

Infectious Disease

Infectious disease outbreaks have posed critical challenges in Canada and worldwide. CGEn nodes have deployed their expertise and resources to support essential research into the viral pathogens responsible for these public health emergencies.

Historically, in the 2002-2004 SARS outbreak, the team at Canada's Michael Smith Genome Sciences Centre (now CGEn-Vacouver) was the first to sequence the new SARS coronavirus[62], The Centre for Applied Genomics (now CGEn-Toronto) facilities supported development of SARS diagnostics and testing in Toronto and Hong Kong[63] and McGill Genome Centre's (now CGEn-Montreal) K. Dewar led the team that sequenced the genome of the deadly C. difficile strain that was prevalent in Quebec hospitals in 2004[64].

During the recent COVID-19 pandemic, CGEn led WGS on human host genomes (see "HostSeq" above). In addition, **CGEn-Montreal was the major sequencing partner for Québec's SARS-CoV-2 studies, working in collaboration with the Institut de Santé Publique du Québec** (INSPQ; work led by I. Ragoussis). Funding from CFI supported increases in sample processing capacity via automation and data analysis (I. Ragoussis, G. Bourque), including for real-time sequencing impacting outbreak control and detection of variants-of-concern. These data were used by the INSPQ to confirm and enhance epidemiological data on the pandemic in Quebec and to help guide the responses of public health authorities[65]. During the height of the pandemic, the Deputy Health Minister of Quebec was receiving weekly updates on CGEn virus sequencing activity to assist with policy decisions. The early success of this genomics-based surveillance effort led to the formation of Canada's Variant of Concern monitoring network (CoVaRR-Net; funded by CIHR from 2021-2025), in which I. Ragoussis led the Genomics Pillar, and J. Shapiro and G. Bourque (associated with CGEn-Montreal) the Bioinformatics and Database Pillar. These efforts were extended to include wastewater analysis research, which was presented to governmental committees and further funded by the Quebec Ministry of Health to develop and establish wastewater surveillance methodologies[66] to identify the presence of SARS-CoV-2 in the community[67]. Both short- and long- read sequencing capabilities of CGEn played a key role in these national efforts.

Together, these experiences have informed pandemic preparedness and pathogen monitoring in general, setting a framework for more rapid and comprehensive responses in the future. Further work, expanding the range of pathogens to be surveyed from environmental samples is carried out though funding from the Public Health Agency of Canada and Genome Quebec (I. Ragoussis and McGill University's D. Frigon).

Select Exemplar Research Projects

enabled by CGEn

Conservation, Agriculture and Other Sectors

Return to Contents

Conservation, Agriculture and Other Sectors

Research based on living things is key to understanding the world around us. Many sectors critical for Canadian prosperity, culture and well-being are based on, or impacted by, plants, animals, fungi, bacteria and other forms of life. Twenty-five per cent of CGEn's users over the last 10 years (more than 750 unique research groups) worked primarily in areas outside of human and health studies. CGEn has led, contributed expertise and generated data for large-scale initiatives (CGEn's CanSeq150 and the Canadian BioGenome Project), as well as more focused studies, supporting research at academic institutions, federal and provincial government bodies, private industry, and not-for-profit entities.

750+

species have been sequenced by CGEn

25%

of CGEn users work primarily in areas other than human health

In **biodiversity and conservation research** genomics reveals genetic diversity to help protect species, ecosystems, and support sustainable stewardship.

In **evolutionary biology** genomics explains how species evolve and adapt over time, informing the understanding of life's history, human origins and disease.

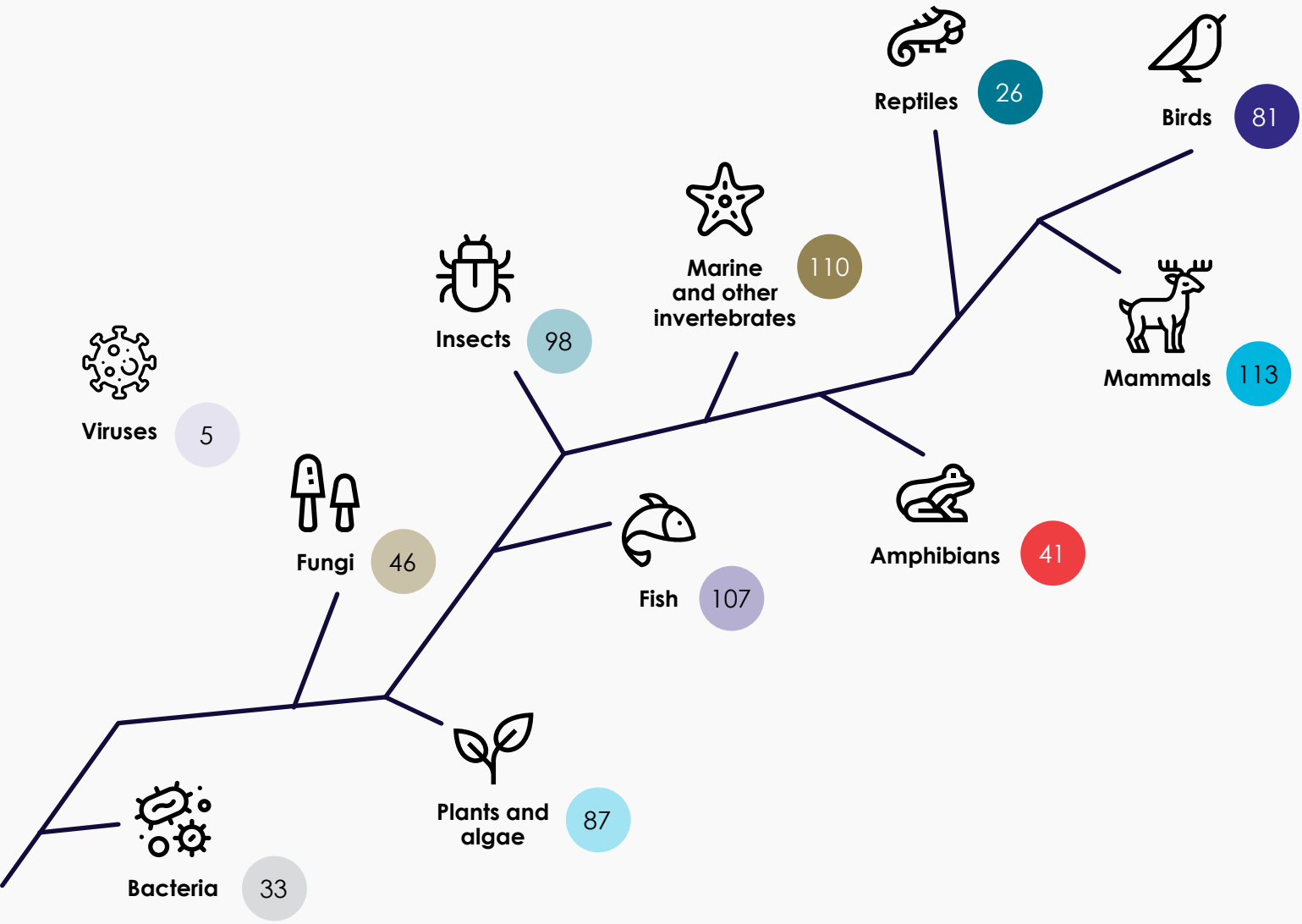
In **commercial sectors** genomics reveals the genetic basis of disease resistance, productivity and resilience across key Canadian industries (agriculture, fisheries, forestry, and more).

Sequencing the diversity of life at CGEn

CGEn has generated high-quality data for more than 750 species of life on Earth. The species sequenced span across all major biological lineages, supporting research across a wide variety of sectors. This includes:

47 **Endangered or threatened species**

156 **Commercially significant species**



Research project examples

Conservation

CGEn-Vancouver generated WGS for 429 **Bull kelp** and 211 **Giant kelp** from the coastlines of British Columbia and Washington. The study, led by J. Bemmels and G. Owens (University of Victoria) and the Kelp Rescue Initiative, aimed to understand evolutionary dynamics and inform restoration strategies of kelp forests. Six to seven geographically and genetically distinct clusters in each species were identified, highlighting the need for targeted regional approaches to conservation[68].

Evolutionary biology

The FITNESS project, led by C. Peichel (Universität Bern) and R. Barrett (McGill University), is the first large-scale, forward-in-time study of parallel evolution in whole ecosystems using **Threespine Stickleback fish**. With over 9,000 out of 11,000 genomes already sequenced by CGEn-Montreal, the project aims to uncover how historical and selective forces shape evolution. The team is also developing a custom genotyping array from the identified variants and will genotype over 10,000 fish in 2026 at CGEn-Montreal.

A. Shafer's team (Trent University) extracted and analyzed DNA from ancient antlers, uncovered while digging the Toronto subway system in 1976, to reconstruct past **Pleistocene-era deer** populations. *Torontoceros hypogaeus* is represented by this single specimen (held at the Royal Ontario Museum) and is dated at approximately 11.3 thousand years ago. Comparison of the *T. hypogaeus* genome, sequenced at CGEn-Toronto, to a range of modern and other ancient deer species (most of which were also sequenced at CGEn-Toronto) revealed where they diverged and areas of the genome that are especially dynamic. This work helps to pinpoint evolution tied to climate or ecological change, and our understanding of how deer species responded differently over time[69].

Forestry

R. Hamelin (University of British Columbia) leads the CoAdapt Tree Project. A recent study used fungal genomic sequencing data generated at CGEn-Montreal to demonstrate that pathogenic **Dutch elm disease, Dothistroma needle blight, and Swiss needle cast** respond similarly to environmental pressures, especially precipitation and humidity, with some genes showing signs of convergent evolution. Understanding these shared adaptive responses under climate change provides insights that could enhance forest disease management and guide sustainable forestry practices[70].

Agriculture

CGEn is supporting the PeaCE (pea climate-efficient) project, led by M. Samuel (University of Calgary) and S. Kagale (National Research Council Canada), which aims to enhance the quality, profitability, and resilience of **Field peas** grown in Canada. CGEn-Vancouver has completed sequencing of over 100 samples using both short- and long-read sequencing platforms, to help reveal the genomic variations underlying climate resilience and root rot resistance traits.

In 2013, I. Birol (associated with CGEn-Vancouver) and C. Helbing (University of Victoria) completed sequencing and *de novo* (for the first time) genome assembly of the **North American bullfrog**, leading to the discovery of novel potential antimicrobial peptides (AMPs)—naturally occurring proteins that help species fight bacterial infections[71]. Since then, Birol and his team have been working to develop alternatives to conventional antibiotics based on AMPs, for use in agricultural practices. The team has developed a high-throughput discovery pipeline, using computational methods to identify novel AMPs from genome sequences and machine learning techniques.

The **Leaf rust fungus** poses a significant threat to wheat crops in Canada and globally, causing yield losses and impacting food security. G. Bakkeren's team at Agriculture and Agri-Food Canada (Summerland Research and Development centre) leveraged expert consultations and specialized infrastructure at CGEn-Vancouver, including advanced long-read sequencing technologies, to tackle the challenges posed by these fungi. These include finding ways to monitor ever-evolving pathogen populations and understand their pathogenic potential to eventually develop more resilient wheat varieties and sustainable crop protection strategies.

Mining

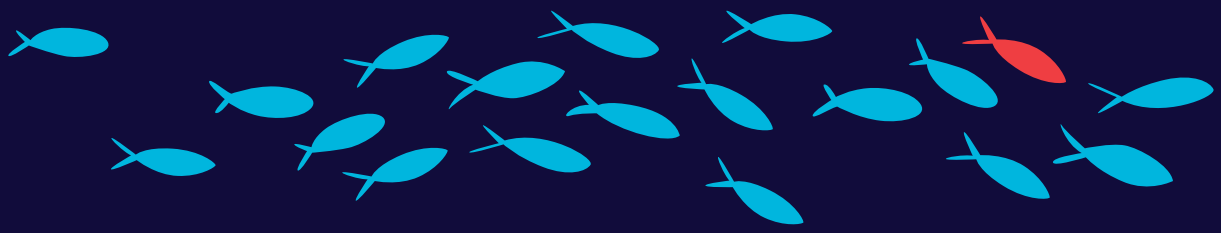
The Mining Microbiome Analytics Platform (M-MAP) is the first integrative platform for collecting and analyzing genomic data from water, soil, and rock in mining environments. The extensive industry-focused databank of genomic and environmental (geospatial, climate and chemical) data about thousands of **microbes** from mining-related sites will help develop cleaner, more efficient mining practices with fewer environmental impacts. CGEn-Vancouver supported the project by generating genomic data for 300 samples included in the platform. Looking to the future, M-MAP has been commercialized into a new startup, nPhyla, which works with clients and partners to convert DNA data into new applications to meet mining challenges.

Sequencing non-human samples can be challenging

Non-human genomes

- May originate from contaminated or limited samples
- May be very large or include multiple sets of chromosomes
- Can include DNA elements that add complexity to sequencing and analysis workflows

Genomic investigations require a blueprint, called a **reference genome**, to guide genome assembly and help researchers identify mutations that explain characteristics, diseases or evolutionary history.



A reference genome is the most complete, highest-quality genome of a species available for research.

If no reference genome is available, it must be assembled without the benefit of a guide, like putting together a jigsaw puzzle without a picture on the box—requiring expertise, a high-quality sample, and advanced computational resources.

CGEn is generating genomic data and supporting the assembly of high-quality reference genomes for hundreds of species.

Supporting research with high-quality reference genomes

As species become increasingly threatened by human activity, research based on reference genomes is advancing biodiversity monitoring, conservation strategies, ecosystem restoration and commercial applications. To support these areas of research, CGEn is an active partner in the **Canadian BioGenome Project**, and led its predecessor **CanSeq150**, generating genomic data that is being assembled into high-quality reference genomes for hundreds of species. Dedicated technology development efforts across CGEn also deliver new, high-quality genome assemblies. These programs advocate for making genomic data available for future research, ensuring investigators can access high-quality genomic resources to support a more sustainable future for Canada and all its inhabitants.

CanSeq150

CGEn launched its CanSeq150 program in 2017 in honour of Canada's 150th birthday, with the objective of providing genomic data to support future biodiversity and conservation research as well as fuel technology development. CanSeq150 was led by S. Jones (CGEn-Vancouver), S. Scherer (CGEn-Toronto), and N. Aziz (formerly CGEn CEO), with I. Ragoussis (CGEn-Montreal), M. Engstrom (Royal Ontario Museum) and L. Bernatchez (Université Laval). In collaboration with academic, government, and NGO organizations, CGEn nodes generated short-read and long-read sequencing data for 95 species with cultural, economic, and ecological importance to Canada that were lacking reference genomes. These data were returned to CanSeq150 project investigators for genome assembly and further analysis, which often involved CGEn informatics experts.

Dr. Mona Nemer, Canada's Chief Science Advisor, speaking at the launch of CanSeq150

For many CanSeq150 species, collaborations were forged for species prioritization, sample acquisition, to formulate scientific questions and bring additional expertise into CGEn—including international partners who added scientific value or could assist for species found across borders. The CanSeq150 program promoted public release of genomic data to enable future research and to date, resulted in 73 publications (associated with data from 23 species) from across the globe.

Canadian BioGenome Project (CBP)

Launched in 2021, the Genome Canada-funded Canadian BioGenome Project (CBP) is led by CGEn-Vancouver Scientific Director S. Jones and M. Murray (University of Calgary) and co-led by I. Ragoussis (CGEn-Montreal), CGEn-Toronto Scientific Director S. Scherer and other leading Canadian researchers. CBP took forward the momentum of CGEn's CanSeq150 initiative, expanding Canadian efforts to comprehensively characterize the genomes of Canadian species. Under the auspices of the CBP, CGEn-Vancouver is part of the Earth BioGenome Secretariat, placing CGEn in a leadership position internationally and providing connections to important programs worldwide. CBP's ambitious effort will deliver 400 high-quality reference genomes for species relevant to Canadian conservation planning, culture, biodiversity, and the economy. DNA from CBP samples is extracted and sequenced at one of CGEn's three nodes, integrating both long- and short-read sequencing (RNA and DNA) technologies. HiC sequencing technology is utilized to uncover the 3D organization of a genome. The resulting data are assembled into reference genomes at CGEn-Vancouver, and fully-annotated (where genes and functional elements in the genome are "labeled" to facilitate research) in partnership with the European Molecular Biology Laboratory – European Bioinformatics Institute (EMBL-EBI). At each stage, data are immediately made accessible to researchers. CBP species have been selected in collaboration with cultural (including those representing Indigenous Peoples), conservation and commercial group partners, or through the CBP's extensive network of 35 taxonomic committees (involving almost 600 experts). CBP

had received samples from over 748 species as of February 2026, and CGEn nodes had generated 223 long-read sequences, 134 HiC sequences, and 73 sets of RNA sequences. The project has assembled 195 genomes, and 51 of these have been annotated. This includes several commercially important species, such as Atlantic cod, Northern propeller clam, and Jonah crab, as well as the Northern red-backed vole, the ant Formica aserva, the endangered great white shark and critically endangered Vancouver Island marmot. To ensure that partners are equipped to benefit from these reference genomes, CBP is developing policy recommendations on how to responsibly implement genomic tools in wildlife conservation, ecosystem-based co-management, species restoration, and environmental monitoring. The project is also developing a knowledge mobilization platform to bridge the gap between genomics and local communities. As CBP moves toward completing its first 400 reference genomes, it is not only building a foundation for genomic conservation in Canada but also shaping a global model for how science, technology, and community partnership can work together to protect biodiversity.

CanSeq150 and **CBP** include species with challenging sample types and genomic structures, and technology development by CGEn's expert staff is helping to produce research-ready data.

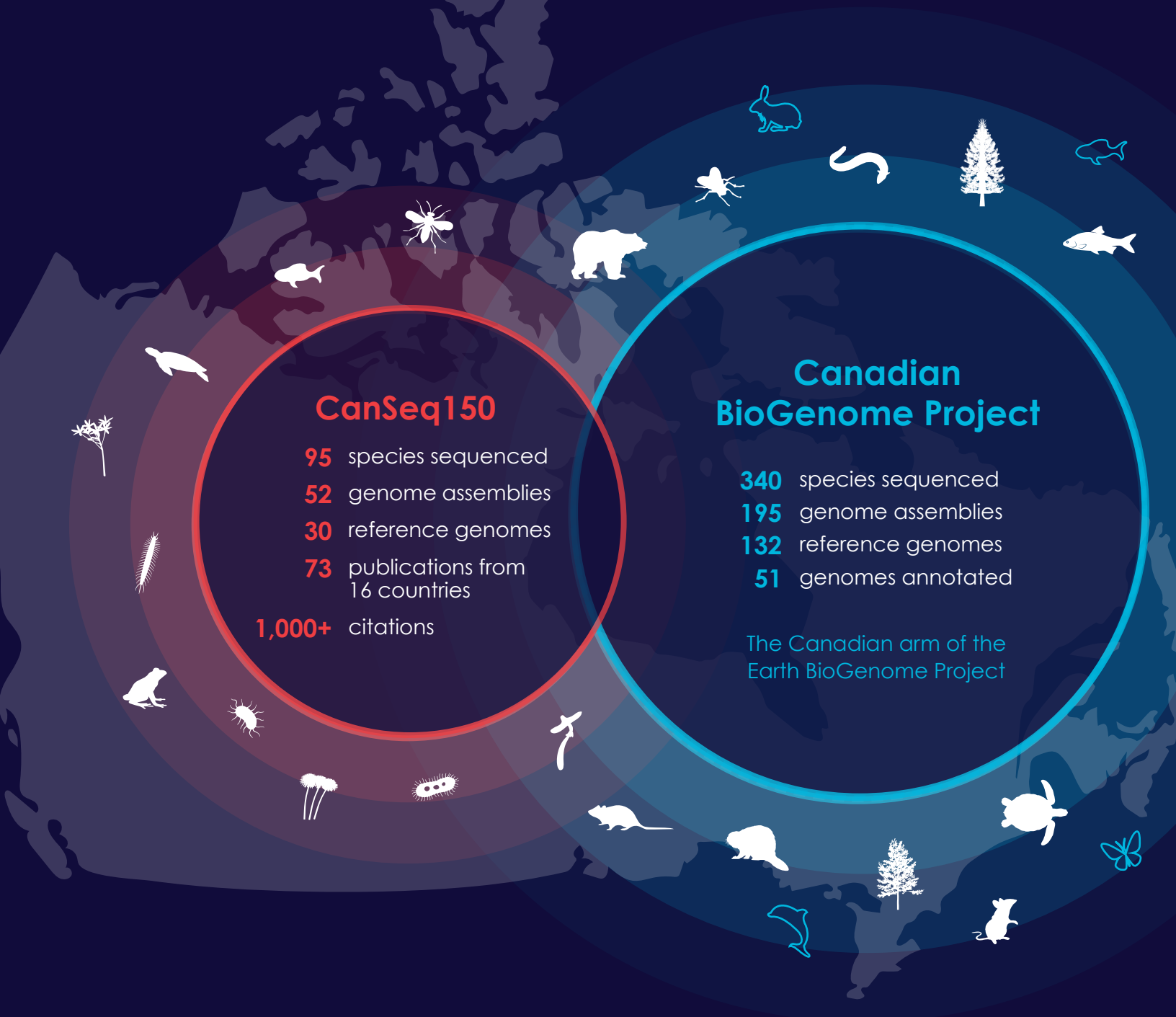
Collaboration between CGEn's three nodes have resulted in development of several methods. For example:

Extraction and clean-up methods have been developed in the context of CBP to produce high-quality DNA samples needed for sequencing, specifically for challenging samples from marine organisms, insects, and plants.

A protocol for generating high-quality sequence information for samples with a very small amount of starting DNA. The Strawberry blossom weevil, an invasive agricultural pest, was the first organism successfully sequenced by CGEn-Toronto for CBP with this method.

Sequencing sample preparation methods, including shearing (breaking up DNA into appropriate sized pieces for sequencing) and size selection, have been optimized for different sample types for long-read sequencing at CGEn-Toronto and CGEn-Montreal.

CanSeq150 and CBP reference data informing research for ecological, evolutionary and economic impact



Examples of research enabled by CanSeq150 and CBP reference genomes

Canada beaver

A study by a Scottish and Norwegian team explored reintroduction of Eurasian beavers to Scotland based on populations from different genetic sources, highlighting the importance of genetic monitoring to guide conservation translocations[72].

An international team showed that Eurasian beavers' specialized chemical sensory system allows them to communicate (including sensing substances floating on water) providing an understanding of social behaviour, territoriality, reproduction, and navigation in aquatic habitats for this ecologically important species[73].

Grizzly bear

The CanSeq150 project genome informed several studies investigating hibernation, beginning with an American project that identified major changes in gene activity across different tissues, suggesting a common regulatory mechanism. As hibernation is often associated with disease-like states, this could inform the development of novel therapeutics to treat human and animal diseases[74].

A Swedish study showed that dental calculus (hardened plaque) can preserve DNA from microbes, hosts, and diet across many mammals, including bears, making it a powerful tool for studying oral health, microbiome evolution, and environmental impacts over time[75].

North Atlantic blue whale

The findings of the CanSeq150 project team revealed genetic resilience to past whaling and evidence of interbreeding with fin whales, offering new insights to guide blue whale conservation in a changing climate[76].

Seaweed fly

The CanSeq150 project team found that large chromosomal inversions play a key role in population responses to changes in climate, salt levels, and habitat, showing that how DNA is reshuffled during reproduction shapes a species' ability to thrive across diverse conditions[77].

A German team showed that kelp fly larvae have a very simple microbiome dominated by just a few bacterial groups, suggesting microbes may play an important role in helping larvae survive on their specialized seaweed diet[78].

Loggerhead sea turtle

Spanish researchers leveraged the Loggerhead genome to investigate the rise in Loggerhead nesting in Spain[80].

American eel

The CanSeq150 project team sequenced 460 individuals, confirming a very similar genetic makeup across their broad geographic range and suggesting that a single, unified conservation strategy may be effective[81].

Data was used by Canadian researchers to help show that genetic markers used in population studies can be misleading because of complex regions in genomes, and that removing these problematic markers makes genetic analyses much more accurate[82].

Greenland halibut

CanSeq150 project researchers sequenced 198 individual halibut to discover that males have a typical XY sex determination system[83]. This study and international investigations involving additional species[84,85], are offering new clues about how sex chromosomes evolve in fish.

A study from the USA showed that different fish species independently evolved nearly identical antifreeze proteins from distinct genetic origins, revealing multiple evolutionary pathways for how entirely new genes and functions can arise[86].

Shrew

The CanSeq150 project team led a comparative genomic study which uncovered shrew-specific genomic variants in genes associated with the nervous, metabolic, and auditory systems. This research is providing insights into shrew adaptation and divergence over time[87].

The team used the data to help show that island shrews in Nova Scotia are smaller, age faster, and show distinct DNA methylation patterns compared to mainland shrews, suggesting that these changes may help drive island environment adaptation[88].

Black spruce

The CanSeq150 project team found key genes involved in stress responses, which will help the forest research community better understand how trees adapt and support the development of molecular-based breeding strategies for this commercially valuable tree[89].

The project team also assembled the complete chloroplast (plant cell structures that carry out photosynthesis) genome of black spruce, providing a key resource for understanding spruce evolution and conifer adaptation[90].

Pronghorn

An international team led by Chinese researchers sequenced reindeer genomes and compared them with 16 other reference genomes, including the Pronghorn, uncovering evidence of multiple domestication events and genetic changes linked to docile behaviour. These findings deepen our understanding of domestication processes and highlight the complexity of related genetic adaptations[91].

The Next 10 Years

Return to Contents

Canadian Genomic Science for National Impact and Global Reach

Over the past decade, CGEn has supported not only the technology and research that drive genomics forward in Canada, but also the people, knowledge, and practices that sustain it. Our efforts have focused on keeping pace with large-scale biology and its growing applications—from genome sequencing and data analysis to data sharing, ethics, and the societal dimensions of genomic science. Working alongside partners across the country, over the next 10 years, we will continue to ensure that Canadian researchers, institutions, and collaborators can benefit from emerging genomic technologies and large-scale data, building capacity for continued global coordination.

Ecosystem Impact

Supporting the Next Generation of Genomics Leaders

To remain current with evolving genomic technologies implemented across CGEn, staff participate in many relevant training activities including rigorous validation and set-up processes, coupled with appropriate vendor equipment-based and technical training. CGEn's distributed structure enables learning and skill-development across nodes, exemplified by benchmarking and technology development collaboration in the Technical Experts Committee.

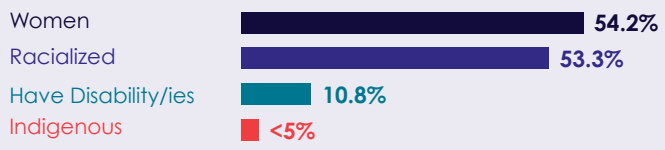
Over the past 10 years, we estimate that CGEn data has supported the research of close to 10,000 highly-qualified research personnel across the country. Enabling Canadian talent with expert advice and high-quality genomic data helps to ensure trainees and research staff develop genomic skills that will be critical for maintaining Canada's world-renowned research enterprise, the innovation economy and global competitiveness. CGEn is committed to working with its users and more broadly with researchers in Canada, to ensure our country is prepared with a well-trained talent pipeline to seize future opportunities in genomic science. Specific examples include CGEn leveraging partnerships with the Canadian Institutes of Health Research and the Canadian Statistical Sciences Institute (Ontario) to support trainees in using HostSeq data.

Equally important is ensuring staff and partners reflect the rich diversity of the research communities CGEn supports. By embracing a wide spectrum of identities, experiences, and knowledge, and endorsing equity as a core value, CGEn strives to promote inclusion as a key driver of excellence in scientific outputs and outcomes. Our Inclusion, Diversity, Equity & Accessibility (IDEA) Strategic Action Plan was developed in 2023, based on relevant literature and resources, best practices at CGEn node host institutions, and the results of a CGEn-wide staff survey which included self-identification and workplace experience questions.

Implementation of the IDEA Strategic Action Plan is overseen by CGEn's IDEA Committee and includes development and delivery of various initiatives to support staff understanding of IDEA in research and in the workplace, including:

- IDEA-focused education sessions (Introduction to IDEA, Sex and Gender in Science and the Workplace, and Implicit Bias) held across CGEn nodes.
- Design of a flexible Mentorship Program Framework used to launch a mentorship program at CGEn-Vancouver in 2025 that supported 14 mentees/mentor relationships. Mentorship topics included women and racialized groups in scientific leadership, developing database skills, and management skills. This pilot will inform future iterations of the program.

How CGEn Staff Self-Identify*



* Select results based on CGEn's Spring 2023 IDEA survey, which received 120 responses (50.6% response rate).

Ecosystem Impact

Supporting the Next Generation of Genomics Leaders

CGEn is committed to advancing public awareness of genomics and its role in health, the environment, and society. Through partnerships with educational and community programs, CGEn plays a role in making complex science accessible and inspiring for all audiences. In 2024, CGEn-Vancouver joined Genome BC to deliver *Geneskool*, introducing high school students across British Columbia to the fundamentals of genomics, its real-world applications, and career opportunities in science and technology. CGEn-Toronto has an ongoing engagement with the *Kids Science* program to spark curiosity through interactive lab experiences. CGEn also collaborated with John Polanyi Collegiate Institute in Toronto for the *CanSeq150 High School Challenge*, where students explored genome sequencing and biodiversity through a proposal-based competition.

Beyond classrooms, CGEn promotes dialogue between scientists and the public on the broader societal dimensions of genomics. For example, the CIHR-funded Café Scientifique event, "*Breaking Down Barriers: Exploring the Role of Equity in the Future of Precision Medicine,*" brought together researchers, patients, and policymakers for a national conversation on equity in genomic health. Together, these initiatives demonstrate CGEn's commitment to building public understanding, trust, and inclusion as genomics continues to shape Canada's future.

As CGEn marks its 10th anniversary, we move into our next phase with purpose and renewed vision.

Over the past decade, CGEn has built a strong foundation of excellence in genome sequencing and analysis that connects researchers, institutions, and sectors across Canada. Through our new Strategic Vision 2030, we will build on and strengthen this groundwork.

Rapid advances in sequencing, informatics, and AI will continue to redefine what's possible in genomics. With continued focus on technology development and training, CGEn will remain Canada's trusted national testbed and hub of genomic expertise, translating innovation into enhanced services that power cutting-edge research and deliver accessible, impactful tools for the scientific community.

CGEn will build on its collaborative model and the spirit of coordination that is gaining momentum across Canada and amongst many countries around the world. Maintaining our existing, and establishing new national and international partnerships that benefit Canadian research, will be a renewed area of focus. Leveraging these connections will support Canadian genomic sciences at large—within the CGEn platform and reaching into research initiatives it supports—enabling Canadian leadership for impact at home and around the world.

CGEn will continue to advance nationally based, large-scale genomics initiatives that are increasingly important in the current global context, leveraging our strengths in high-quality data generation, analysis, and governance. The success of CGEn-led and CGEn-partnered initiatives such as HostSeq, the Marathon of Hope Cancer Centres Network, CanSeq150, and the Canadian BioGenome Project, amongst many others, has firmly demonstrated the importance of Canada's national platform for genome sequencing and analysis. Looking to the future, CGEn will support a new pipeline of projects, including Genome Canada's 100,000-genome Canadian Precision Health Initiative, that will have lasting impacts in research and its downstream applications.

CGEn's Strategic Vision 2030 builds on our core strengths—scientific excellence, collaboration, agility, and inclusivity—to advance world-class genomics in Canada. By investing in people, partnerships, and innovation, CGEn will continue to power discovery across health, environment, agriculture, and many more sectors, shaping a healthier, more sustainable, and more prosperous future for all.

Acknowledgements

We extend our sincere thanks to the many individuals, partners, and funders whose leadership, expertise, and commitment make our work possible. Together, CGEn's scientific leadership and staff, Board of Directors, Scientific Advisory Board, and many collaborators continue to guide and strengthen our impact across Canada's genomics landscape.

Board of Directors

Gordon McCauley (Chair)
President & CEO, adMare BioInnovations

Dominique Bérubé
Vice President (Research and Innovation), McGill University

Sir Mark Caulfield
Chief Executive of Barts Life Sciences, Director of the NIHR Biomedical Research Centre at Barts, William Harvey Research Institute, Queen Mary University of London

Philippe Gros
Chief Scientific Officer of the DNA to RNA (D2R) Initiative and Professor, Department of Biochemistry, McGill University

David Malkin
Director, Cancer Genetics Program and Senior Scientist, Genetics & Genome Biology, The Hospital for Sick Children

Sharon Peacock
Master of Churchill College, Professor of Microbiology and Public Health at the University of Cambridge

Vardit Ravitsky
President and CEO, The Hastings Center

Sharon Wu
Managing Director, Vice-President Research & Innovation Portfolio, University of British Columbia

International Scientific Advisory Board

Edward M Rubin (Chair)
Director and Founder, Science Corps, USA

Mark Blaxter
Head, Tree of Life Programme, Wellcome Sanger Institute, Cambridge, UK

Ivo G Gut
Director, Centro Nacional de Análisis Genómico (CNAG), Spain

Elaine R Mardis
Co-Executive Director of the Steve and Cindy Rasmussen Institute for Genomic Medicine at Nationwide Children's Hospital and Professor of Pediatrics at The Ohio State University College of Medicine, USA

Gabor Marth
Professor of Human Genetics, University of Utah, USA

Richard M Myers
Chief Scientific Officer, President Emeritus and Faculty Investigator, HudsonAlpha Institute for Biotechnology, USA

Beth Shapiro
Professor, Ecology & Evolutionary Biology Department Genomics Institute, University of California Santa Cruz and Howard Hughes Medical Institute Professor, USA

Alexandra-Chloé Villani
Director, Massachusetts General Hospital (MGH) Single Cell Genomics Research, and Program Institute Member, Broad Institute of MIT and Harvard, USA

Acknowledgements

Institutional and Funding Partners

SickKids®

Provincial Health Services Authority

UBC THE UNIVERSITY OF BRITISH COLUMBIA

McGill

INNOVATION Canada Foundation for Innovation Fondation canadienne pour l'innovation

GenomeCanada

Ontario

BRITISH COLUMBIA

Québec

Innovation, Science and Economic Development Canada

References

1. Iafrate et al. Nat. Genet. 36, 949–951 (2004)
2. Reuter et al. CMAJ 190, E126–E136 (2018)
3. Marra et al. Science 300, 1399-404 (2003)
4. Jones et al. Genome Biol. 11, R82 (2010)
5. Shah et al. Nature 486, 395–399 (2012)
6. Corbett et al. Front. Genet. 11, 612515 (2020)
7. Haile et al. Front. Genet. 12, 665888 (2021)
8. O'Flanagan et al. Genome Biol. 20, 210 (2019)
9. Maitra et al. Nat Protoc. 16, 2788-2801 (2021)
10. Nagy et al. Nat Neurosci. 23, 771-781 (2020)
11. Sanchez-Ferras et al. Nat Commun. 12, 2627 (2021)
12. Strasser et al. Dev Cell. 60, 2807-2824 (2025)
13. Moore et al. J. Mol. Diagn. 22, 141-146 (2020)
14. Lok et al. G3 (Bethesda) 7, 755-773 (2017)
15. Lok et al. G3 (Bethesda) 12, jkac138 (2022)
16. Lok et al. Sci. Rep. 14, 21023 (2024)
17. Ward et al. Nat Commun. 12, 491 (2021)
18. Meccariello et al. Science. 365, 1457-1460 (2019)
19. Bayega et al. Sci Rep. 11, 7878 (2021)
20. Bayega et al. BMC Genomics. 21, 259 (2020)
21. Krzywinski et al. Genome Res. 19, 1639-1645 (2009)
22. Panjwani et al. PLoS Comput. Biol. 16, e1008336 (2020)
23. Trost et al. Am. J. Hum. Genet. 102, 142-155 (2018)
24. Akbari et al. Genome Biol. 22, 68 (2021)
25. Baskurt et al. Bioinformatics. 36, 1283-1285 (2020)
26. Reisle et al. Bioinformatics. 35, 515-517 (2019)
27. Yoo et al. BMC Genomics 24, 26 (2023)
28. Garg et al. PLoS Genet. 20, e1011192 (2024)
29. Warren et al. HLA 103, e15355 (2024)
30. Mendes et al. Am. J. Hum. Genet. 112, 135-153 (2024)
31. Lin et al. Genome Med. 12, 115 (2020)
32. Povysil et al. J. Clin. Invest. 131, e147834 (2021)
33. Taher et al. BMJ Open 11, e052842 (2021)
34. Tremblay et al. PLoS One 16, e0245031 (2021)
35. The COVID-19 Host Genetics Initiative. Nature 621, E7-E26 (2023)
36. Longstaff et al. CMAJ 197, E941-E952 (2022)
37. Forgetta et al. BMJ Open 12, e059021 (2022)
38. Han et al. Am. J. Hum. Genet. 108, 1204-1216 (2021)
39. Forgetta et al. PLoS Med. 17, e1003152 (2020)
40. Menniti et al. Front. Cardiovasc. Med. 8, 599671 (2021)
41. Laskin et al. Cold Spring Harb. Mol. Case Stud. 1, a000570 (2015)
42. O'Neill et al. Cell Genomics 4, 100674 (2024)
43. Pleasance et al. Nat. Cancer 1, 452-468 (2020)
44. Jones et al. Ann. Oncol. 28, 3092-3097 (2017)
45. Reisle et al. Bioinformatics 35, 515-517 (2019)
46. Jones et al. Clin. Cancer Res. 25, 4674-4681 (2019)
47. Villani et al. Nat. Cancer 4, 203-221 (2023)
48. Deyell et al. Nat. Commun. 15, 4165 (2024)
49. Terry Fox Research Institute Marathon of Hope Cancer Centres Network. Cancer Cell 43, 587–592 (2025)
50. Yuen et al. Nat. Med. 21, 185-191 (2015)
51. Tammimies et al. JAMA 314, 895-903 (2015)
52. Yuen et al. Nat. Neurosci. 20, 62-611 (2017)
53. Schaaf et al. Nat. Rev. Genet. 21, 367-376 (2020)
54. Trost et al. Cell 185, 4409-4427 (2022)
55. Scala et al. Am. J. Hum. Genet. 112, 154-167 (2025)
56. Selvanayagam et al. J. Med. Genet. 62, 413-421 (2025)
57. Fehlings et al. Nat. Genet. 56, 585-594 (2024)
58. Grewal et al. JAMA Netw Open. 2, e192597 (2019)
59. Comitani et al. Nat. Med. 29, 656-666 (2023)
60. Boycott et al. Am. J. Hum. Genet. 109, 1947-1959 (2022)
61. Mitina et al. Lancet eBioMedicine 101, 105027 (2024)
62. Marra et al. Science 300, 1399-1404 (2003)
63. Poutanen et al. N. Engl. J. Med. 348, 1995-2005 (2003)
64. Loo et al. N. Engl. J. Med. 353, 2442-2449 (2005)
65. Murall et al. Genome Med. 13, 169 (2021)
66. Bayega et al. Front Genet. 16, 1516791 (2025)
67. Jayme et al. Viruses 16, 1495 (2024)
68. Bemmels et al. Curr. Biol. 35, 688-698.e8 (2025)
69. Kessler et al. bioRxiv (2025)
70. Hessenauer et al. Genome Biol. Evol. 17, evaf069 (2025)
71. Birol et al. PLoS One 10, e0130720 (2015)
72. Taylor et al. Evol Appl. 17, e13629 (2023)
73. Tomiyasu et al. J Anat. 241, 809-819 (2022)
74. Jansen et al. Commun Biol. 2, 336 (2019)
75. Brealey et al. Mol Biol Evol. 37, 3003-3022 (2020)
76. Jossey et al. Conserv. Genet. 25, 357-371 (2024)
77. Mérot et al. Mol. Biol. Evol. 38, 3953-3971 (2021)
78. Bischof et al. Microb Ecol. 87, 91 (2024)
79. Chang et al. F1000Res 12, 336 (2023)
80. Luna-Ortiz et al. Sci. Rep. 14, 1506 (2024)
81. Ulmo-Diaz et al. Evol. Appl. 16, 1872-1888 (2023)
82. Dallaire et al. Genome Biol. Evol. 15, evad229 (2023)
83. Ferchaud et al. G3 (Bethesda) 12, jkab376 (2022)
84. Jasonowicz et al. Mol Ecol Resour. 22, 2685-2700 (2022)
85. Edvardsen et al. PLoS Genet. 18, e1010011 (2022)
86. Rives et al. Mol Biol Evol. 41, msae182 (2024)
87. Cossette et al. Genome Biol. Evol. 16, evae225 (2024)
88. Cossette et al. Mol Ecol. 32, 152-166 (2023)
89. Lo et al. G3 (Bethesda) 14, jkad247 (2024)
90. Lo et al. Microbiol Resour Announc. 9, e00877-20 (2020)
91. Wu et al. Evol. Lett. 8, 494–504 (2024)

www.cgen.ca
www.linkedin.com/company/cgen