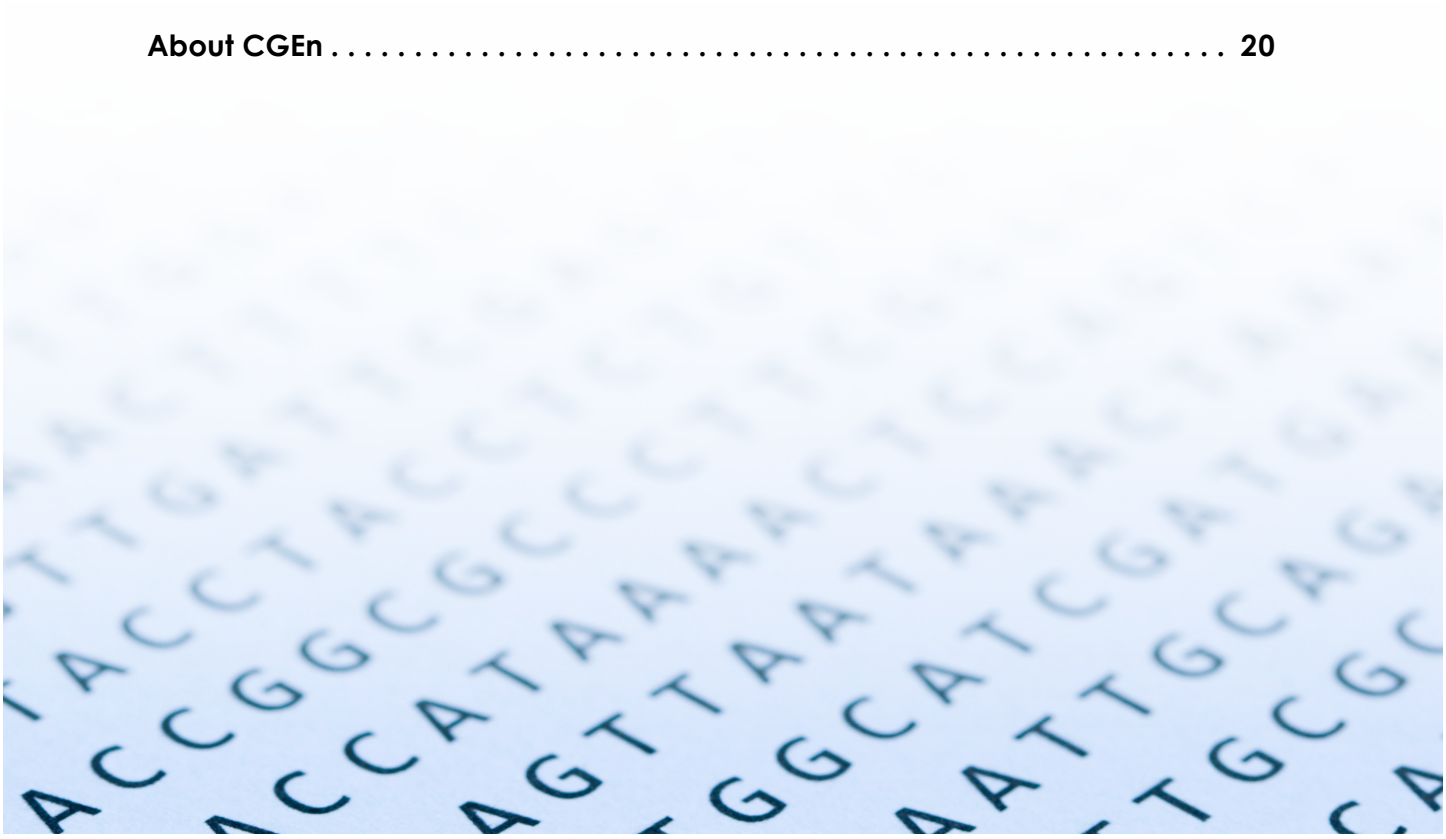
CGEn

Canada's national platform for genome sequencing and analysis

Annual Report
2021-2022

Table of Contents



CEO Message

CGEn provides the leading genomic infrastructure for Canada's research and innovation communities. Since its inception, CGEn has facilitated the sequencing of thousands of human and other species' genomes, catalyzing groundbreaking new research in Canada.

This year, more than 1,000 research groups across Canada accessed our cutting-edge technologies and large-scale data-generation platforms to perform research related to human health, agriculture, forestry, and other areas of importance to Canada and Canadians. The growing need for large scale genomics data to support bioscience research activities across the country is evident given that CGEn produced 60% more data than previous years on average.

Our annual survey of principal investigators supported by CGEn showed that users of our national facilities across Canada are highly satisfied with the overall quality of our services and that our infrastructure and expertise are highly valued. Thanks to the commitment of our advisory and governance boards, funders, partners and highly skilled staff, CGEn will continue to provide essential infrastructure and expertise to support, sustain and enrich genomics research in Canada.

Working together with our partners, we will continue to position Canada at the forefront of genomics research in a long-term, sustainable manner.

Naveed Aziz
Chief Executive Officer, CGEn

Vision & Mission

Our Vision

Serve as Canada's engine for genomics-enabled research and discovery, supporting a healthier and more sustainable future for all Canadians.

Our Mission

Enhance Canada's national capacity for genome sequencing and informatics analysis; **Accelerate** next-generation scientific solutions underpinned by large-scale data generation, and; **Support** Canadian national and international projects in sequencing, databasing and open science collaborations.

Executive Summary

This annual report reflects CGEn's activities and achievements over fiscal year 2021-2022. Despite the continuing challenges presented by the COVID-19 pandemic, CGEn remained fully operational over this timeframe to support the Canadian and international research communities with high-quality genomic services and expertise while also leading national-scope genomics projects.

This past year, CGEn engaged over 1,000 individual labs from across Canada and beyond, supporting projects large and small, and producing over 2.3 petabytes of data – a 60% increase over the average output of the previous four years. CGEn's annual user survey indicated that our user base is highly satisfied with the overall quality and value of CGEn's services – and that CGEn provides essential infrastructure and expertise to sustain genomics research in Canada.

CGEn continues to lead and support multitudes of research projects across many areas of study, several of which are featured within this report. CGEn continued its active leadership of the HostSeq initiative that is generating genomic sequence information and collecting matched clinical data for 10,000 people affected by COVID-19 in Canada. The resulting HostSeq Databank is now accessible to researchers from Canada and beyond. CGEn also launched the Canada BioGenome Project which will deliver 400 reference quality, annotated genomes of Canadian species relevant to conservation planning, biodiversity, and the economy. In cancer research, CGEn is supporting sequencing activities of the Terry Fox Marathon of Hope Cancer Centres Network, the ongoing Personalized Onco-Genomics program, PROFYLE program, and many others. CGEn will also continue to support large-scale disease, disorder and population studies including EPIBRAIN, MSSNG, Psychiatric Genomics Consortium, "Spit for Science" amongst many others.

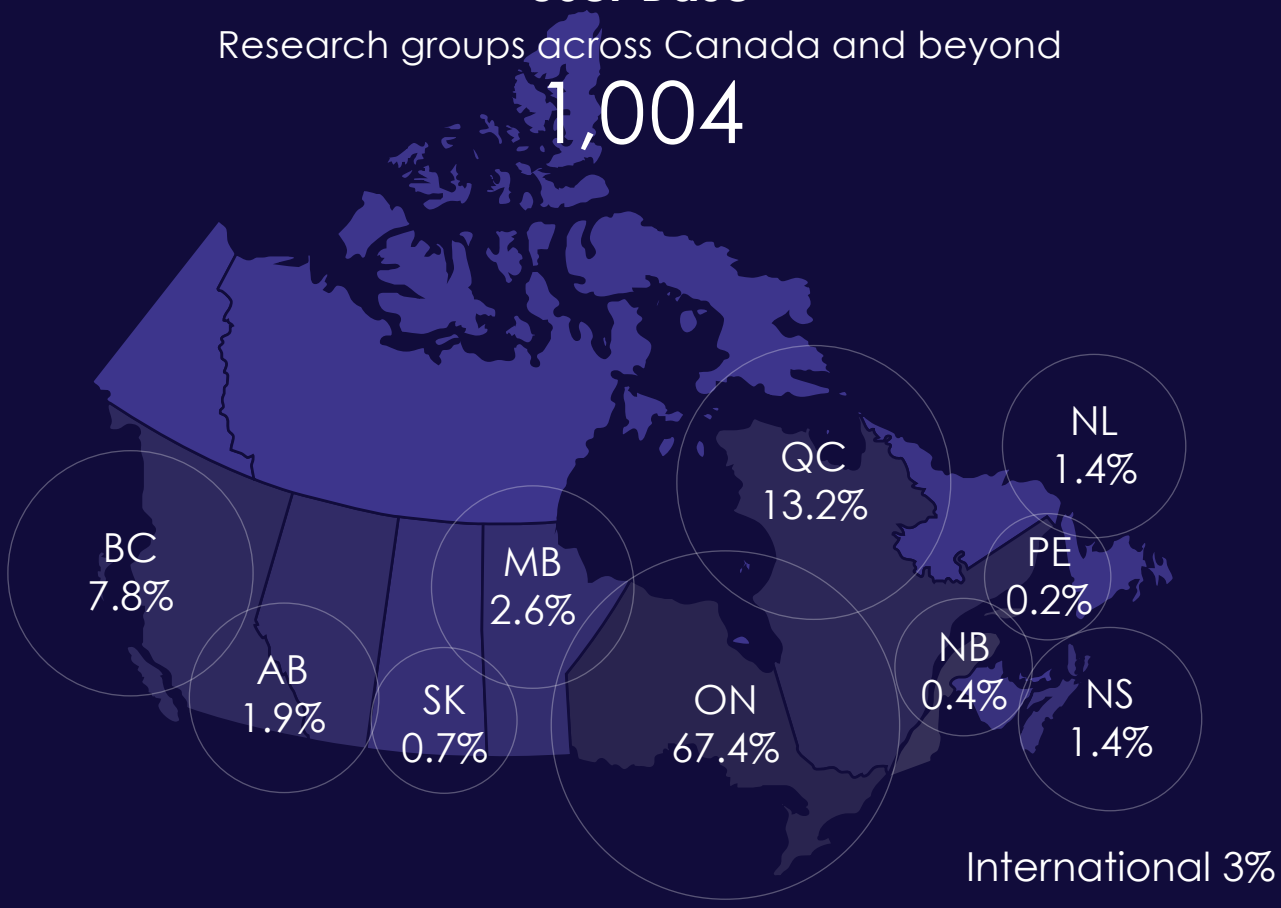
As the demand for genomics grows in Canada's research ecosystem, CGEn will expand capacity and continue to provide the highest quality genomic data to Canadian and international research projects regardless of their size or area of study.

2021-2022 Impact

User Base

Research groups across Canada and beyond

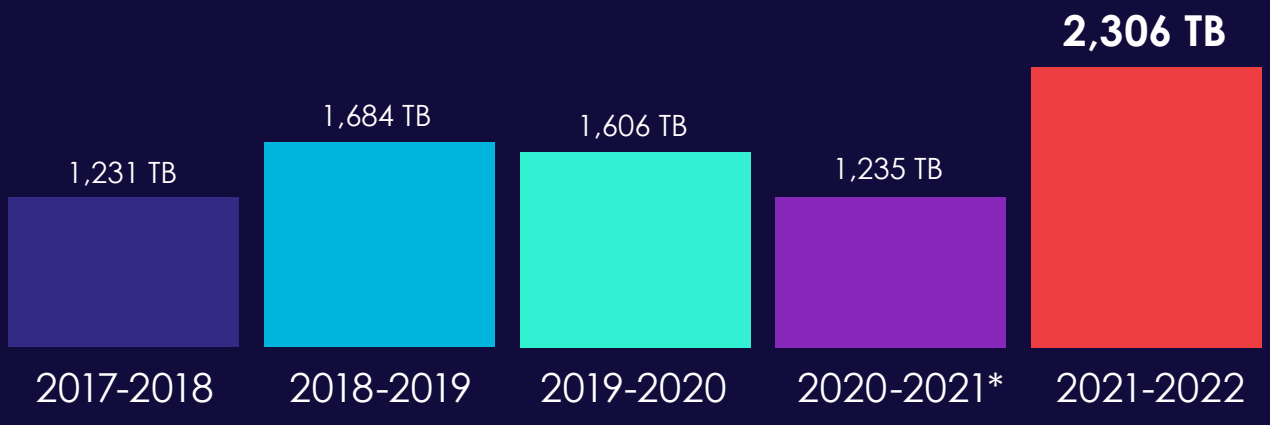
1,004



International 3%

Data Generation

Terabytes of data produced by CGEn



*The 2020/21 drop is largely due to community research activity slowdown resulting from the COVID-19 pandemic.

Training

HQP utilized CGEn-generated data

3,721

57
Undergraduate students

297
Master's students

433
PhD students

599
Postdoctoral Fellows

2,056
Technical and professional

240
CGEn staff – technical

28
CGEn – administrative

11
Other

Advancement of Research

Key knowledge transfer activities by CGEn staff

User Satisfaction

Proportion of responders satisfied

CGEn: Supporting Canadian Science

CGEn's contribution to Canada's COVID-19 response

Implementation Committee:

Drs. Naveed Aziz, Steven Scherer, Lisa Strug, (The Hospital for Sick Children, Toronto), Bartha Knoppers, Mark Lathrop (McGill), Steven Jones (Michael Smith Genome Sciences Centre), and Stuart Turvey (University of British Columbia)

The Host Genome Sequencing Initiative (HostSeq) was launched in July 2020 in response to the COVID-19 pandemic, as part of the Government of Canada's Canadian COVID-19 Genomics Network (CanCOGeN). Supported by Genome Canada, ISED, and CFI-MSI operational funds, CGEn is generating genomic sequence information and collecting matched clinical data for 10,000 people affected by COVID-19 in Canada, with the aims of understanding variable disease outcomes, identifying new biomarkers for risk prediction, and creating a national platform and genomics-related network to prepare for future biological crises.

Over the last year, HostSeq has continued to engage clinical studies to use the CGEn-developed, nation-wide framework for participant recruitment that features research ethics board (REB) protocols and consents with standardized language. HostSeq now includes a network of 14 studies from across Canada, many of which are funded by CFI, CIHR, and specially allocated institutional funds. The protocol enables data sharing broadly as well as use of the data for future research, patient recontact, and data linkage. This standardized, national approach has enabled CGEn to overcome the challenges of provincial division of health care – as of March 31, 2022, CGEn had received 9,652 biological samples that meet the study requirements.

"The emergence of COVID-19 at the footsteps of SARS and MERS highlights a significant issue – that there will be similar outbreaks of severe infectious disease in the future. This pan-Canadian HostSeq initiative addresses the challenges of the COVID-19 pandemic, prepares Canada for a possible re-emergence, and lays the foundation to handle future pandemics."

Dr. Stephen Scherer, CGEn-Toronto Scientific Director, Professor of Genome Sciences, University of Toronto, and Chief of Research, The Hospital for Sick Children (SickKids)

Whole genome sequencing (WGS) of HostSeq samples continues at a minimum of 30X coverage using the Illumina NovaSeq 6000 – samples are processed at one of the three CGEn sites, based on capacity and geographical proximity. At the time of writing this report, a total of 7,068 genomes have been sequenced and are being analyzed for all types of genomic variation, including SNVs, copy number/other structural variation, and mitochondrial variants. If requested, the data is returned to the contributing PIs to enable future research. To support databank-wide, epidemiological studies, genomes are joint called with sample- and variant-level quality control. This includes sample and variant completeness, Hardy-Weinberg equilibrium testing, cryptic relatedness assessment, and ancestry estimation.

"We've created the infrastructure, including recruitment, storage, sequencing and clinical data capture, for a large-scale national whole genome sequence project. HostSeq was designed with a very open consent, where individuals who participate in the study have agreed to share and link their data for any health-related research questions. This resource has great potential to support future health research across Canada beyond this pandemic."

Dr. Lisa Strug, HostSeq Genetic Epidemiology Committee Chair, Professor, University of Toronto, Senior Scientist, The Hospital for Sick Children (SickKids)

Clinical data collection across participating sites is modeled after the WHO ISARIC case report form, and data are harmonized to this standard as they are submitted to HostSeq. The genomic data are linkable with other national, provincial and local data to support analyses, while enabling matching for confounding and intrinsic variables. Aggregate genomic and clinical data are available publicly on the CGEn website. In addition, CGEn scientists, led by Dr. Lisa Strug, recently released a preprint that provides an overview of HostSeq's overall design as well as summary level information from the HostSeq databank.

In parallel to ongoing study recruitment, WGS and clinical data activities, 2021/2022 saw the creation of the centralized HostSeq databank and the implementation of processes for databank access by the research community. The HostSeq Data Access Compliance Office (DACO), led by Ma'n Zawati (McGill), coordinates access requests proposing to utilize the HostSeq databank. Over the past year, 13 studies have completed the HostSeq data access process. This research will help to identify biomarkers for disease severity, predict treatment response, identify genetic risk

and protective factors, and is utilizing and developing novel tools and techniques that support the study of the host genetic components of COVID-19. For example, Dr. Pingzhao Hu (University of Manitoba) will use the HostSeq data to inform the development of AI-based methods to calculate polygenic risk scores of complex traits involved in COVID-19 severity. In another study, Dr. Jordan Lerner-Ellis (Sinai Health) is investigating how individuals' different HLA alleles may result in stronger or weaker immune response against SARS-CoV2, and may lead to accurate targeted treatments, prevention and prophylaxis.

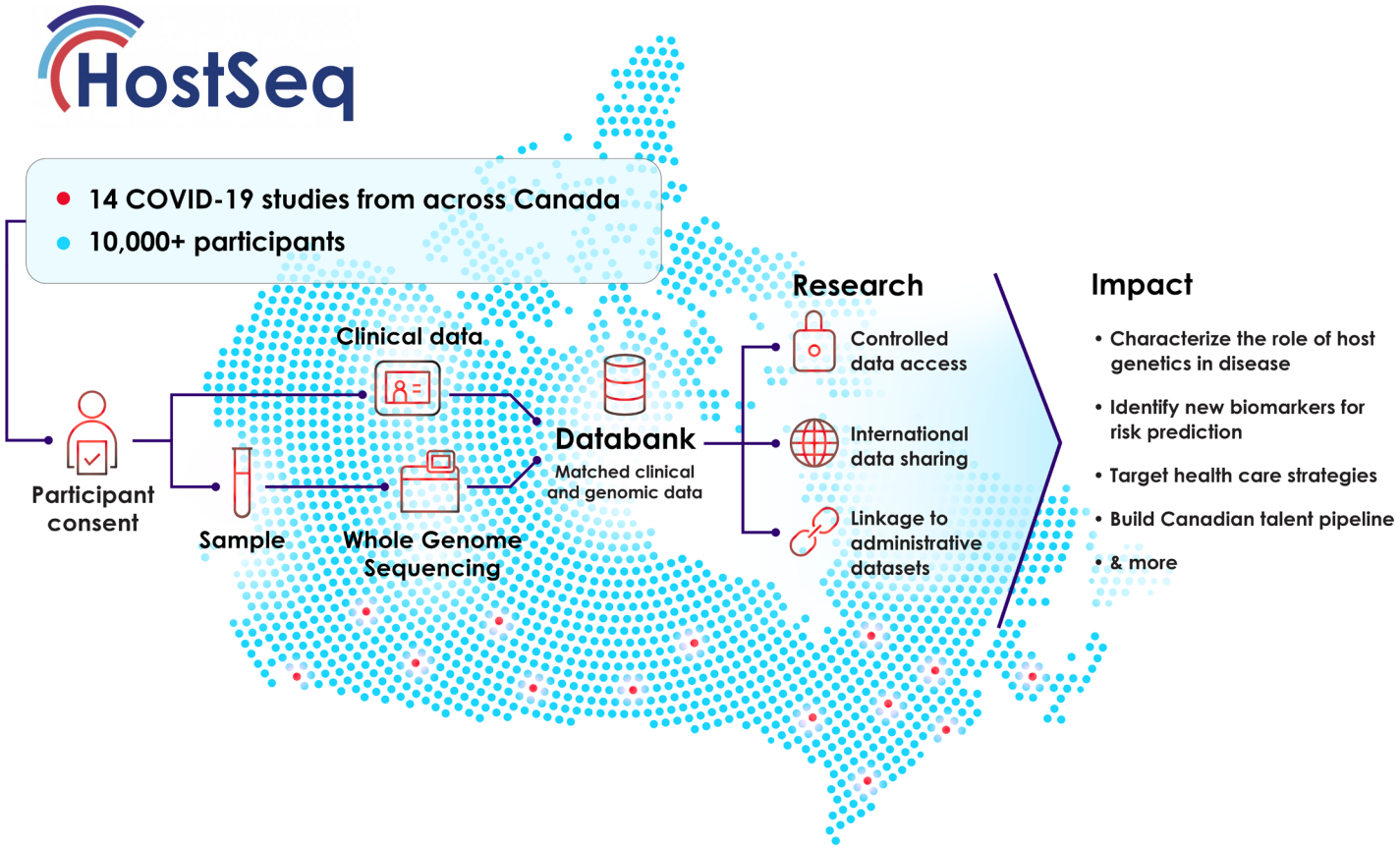
"This project allows us to have this amazing dataset of clinical information about how people were before infection and how they experienced that infection; we have the sequence information about the specific virus that infected them, and then we have the whole genome sequencing. Through that we can put together this remarkable picture about the host, about the virus and about the clinical outcomes."

Dr. Stuart Turvey, HostSeq Study Recruitment Sub-Committee Chair, Investigator, BC Children's Hospital, Canada Research Chair in Pediatric Precision Health

As CGEn works toward depositing 10,000 genome sequences and clinical records for individuals affected by COVID-19 in the HostSeq databank (estimated completion Q4 2022), we have begun to secure partnerships that further promote HostSeq data utilization. For example, three fellowships were awarded in 2021/2022 through a joint CGEn/Canadian Statistical Sciences Institute program aimed at enhancing skills of senior trainees (postdoctoral fellows, graduate students), leveraging the HostSeq databank and CANSSI's expertise. While focused on COVID-19 research, the skills and expertise developed through this program will feed a pipeline of data scientists that will be critical for advancing genomic medicine applications in Canada.

One of the goals of HostSeq is to enable collaborations with international COVID-19 genetics efforts. In 2021 HostSeq contributed data to the COVID-19 Host Genetics Initiative (www.covid19hg.org, Nature 2021). In addition, HostSeq has already contributed to a number of publications addressing study design and genetic variation associated with COVID-19 (Povysil, J Clin Invest 2021 doi: https://doi.org/10.1172/JCI147834; Taher, BMJ Open 2021 doi: https://doi.org/10.1136/bmjopen-2021-052842; Tremblay, PLoS One 2021 doi: https://doi.org/10.1371/journal.pone.0245031; Zhou, Nat Med 2021. https://doi.org/10.1038/s41591-021-01281-1; Nakanishi et al. doi: https://doi.org/10.1172/JCI152386).

The HostSeq databank facilitates sharing with national and international researchers for COVID-19 related research and beyond, and positions Canada to incubate large-scale population genomic sequencing initiatives that can eventually inform diverse health outcomes for all Canadians.



Partners:

Using genomics to take on incurable cancers in children

Principal Investigator:

Dr. David Malkin, Senior Staff Oncologist, Division of Haematology/ Oncology, and Senior Scientist, Genetics and Genome Biology Program, The Hospital for Sick Children (SickKids) in Toronto

"Nuanced differences are very hard to tease out on an individual basis, so the most effective way to do that is by pulling large datasets together, and then to use the large amount of information and apply it to the individual patient in front of you. We could do sequencing without CGEn, but we wouldn't have the capacity to fully interrogate all of this information. So, it's been absolutely key."

About KiCS Program

The SickKids Cancer Sequencing Program (KiCS) is a research study at The Hospital for Sick Children (SickKids) in Toronto designed to learn how to best use a new type of genetic test, called next generation sequencing (NGS) in clinical cancer care. This program is for children who have been diagnosed with a solid tumour, a tumour of the central nervous system (brain tumour) or a blood cancer (leukemia or lymphoma).

The NGS testing is used to learn about the genome of a child's cancer and healthy cells. By doing this, researchers want to know if this new test can help:

1. Better characterize a child's tumour and identify its unique genetic fingerprint.
2. Use this information to identify specific treatment options for a child's tumour.
3. Use this information to follow a tumour's response to treatment.
4. Gain information about a child's prognosis.
5. Understand the cause of a child's cancer.
6. Determine whether there is an underlying genetic predisposition to the cancer.

This study is being led by Drs. David Malkin, Anita Villani and Adam Shlien at SickKids.

About PROFYLE (PRecision Oncology For Young peopLE)

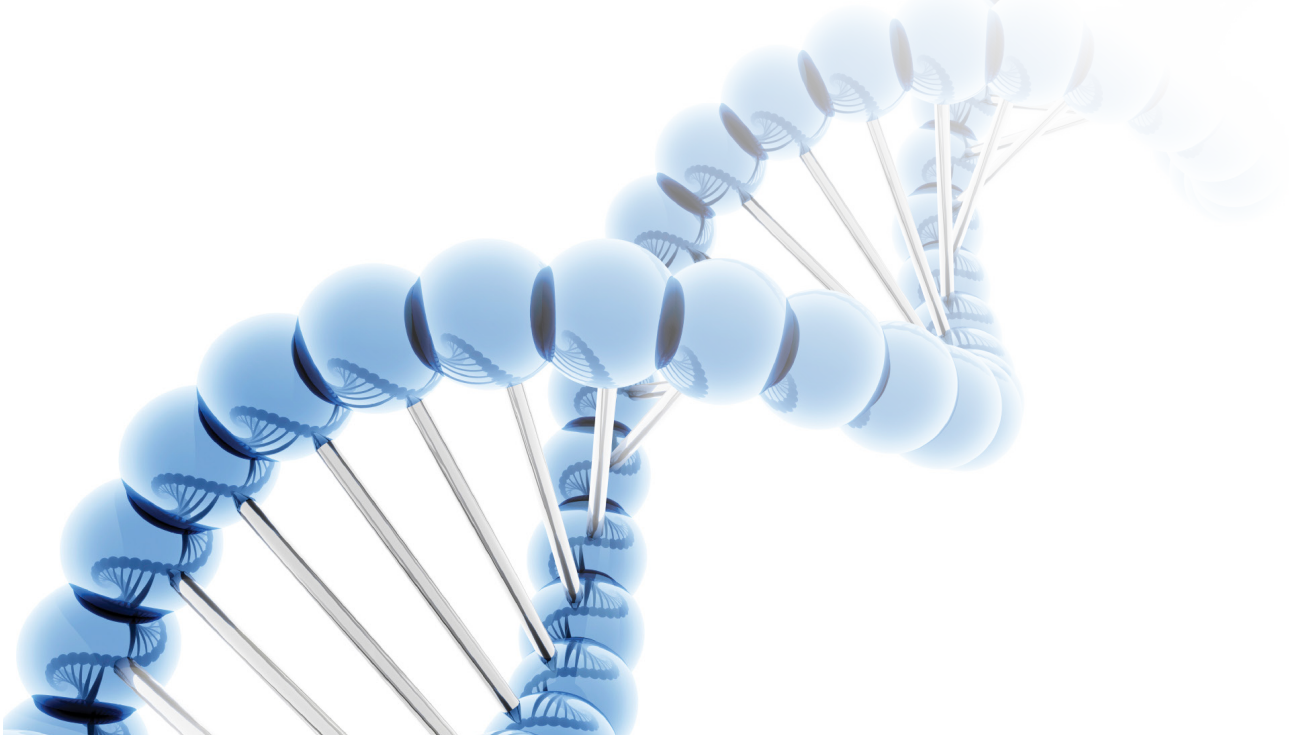
More than 30 paediatric and adult cancer research and funding organizations have joined forces through PROFYLE, a pan-Canadian program to transform the care of children, adolescents and young adults (CAYA) with refractory, relapsed and metastatic ('hard-to-cure') cancer across Canada by using next-generation molecular tools and cancer model systems to identify disease- and patient-specific biomarkers that are tractable targets for therapy. The ultimate goal of the program is to positively impact the lives of CAYA patients across Canada (no matter where they live) with hard-to-cure cancer.

Research centres and top scientists and clinicians across Canada are working together to molecularly profile the tumours of young cancer patients. PROFYLE seeks to take advantage of the genomic revolution to look at the molecular composition of these cancers. Molecular profiling is a set of emerging biological tests that looks at an individual's cancer, studying its genetic characteristics and any unique biomarkers. The information gathered will be used to identify and create therapies that are designed to target a specific cancer tumour profile so patients can enjoy a better quality of life and live longer.

In the past, if a child battling cancer lived in a region without a molecular profiling site, accessing this was not an option. PROFYLE has broken down barriers, giving young people who need it the most access to the best cancer care in Canada.

The PROFYLE Program involves the three sequencing centres that align with CGEn's national platform nodes in Vancouver, Toronto and Montreal. The PROFYLE Program Executive Committee (PEC) is responsible for planning and managing all of PROFYLE program activities. Members include CGEn staff and leadership, including Dr. Steven Jones, Scientific Advisor, CGEn-Vancouver node.

Dr. David Malkin is leading the PROFYLE initiative as Program Director.



Informing childhood brain cancer treatment with genomics

Principal Investigator:

Dr. Nada Jabado, Professor of Pediatrics and Human Genetics at McGill University and a staff physician in the Division of Hematology and Oncology at the Montreal Children's Hospital

"The mutations that we identified not only are giving us a means to better understand and hopefully soon to better treat those children with deadly cancer, but they are amazing tools to tease apart the developmental stages. These mutations can be used as another model system to ask what this gene and protein are doing at a certain time point in development; when is it needed and when it is not needed."

Tackling childhood brain cancer at the root to improve survival and quality of life

Brain cancer remains a lethal and disabling disease, the leading cause of cancer-related deaths among children under age 20 and the third-leading cause in young adults aged 20-39. There are particularly aggressive forms of brain cancer, with barely 10 percent of children and young adults surviving three years after diagnosis, and other forms where those who do survive suffer severe lifelong disabilities due to the life-saving therapies they receive.

Through our previous large-scale genomics projects ICHANGE and MAGIC, our research group and collaborators Drs. Jacek Majewski of McGill University and Michael Taylor at SickKids have discovered that many pediatric brain tumours are driven by mutations in genes that play a significant role in brain development. To decrease the burden of survivorship and improve survival rates, this project is focused on fast tracking the use of treatments targeting specific genetic alterations early at diagnosis. We are also performing innovative investigations of the tumour genome and transcriptome, including at the single-cell level, to identify new alterations and specific vulnerabilities that can be targeted for therapy. Our team will ensure treatments are validated through relevant disease models and fast-track meaningful clinical trials to tackle refractory brain tumours; the goal is to work closely with health-care providers and regulators to ensure the rapid translation of validated treatments to the bedside.

Canada BioGenome Project

Principal Investigator:

Dr. Steven Jones, Scientific Director, CGEn-Vancouver node and Director, Head of Bioinformatics and Distinguished Scientist at Canada's Michael Smith Genome Sciences, BC Cancer

"CGEn has been able to set up sequencing infrastructure across the country, and most of the demand, out of necessity, has been with human health, and so we can sequence human genomes now at scale, and at a population-level basis. One of the things that we wanted to do is not just be involved in human health, but see how we can point CGEn technology and expertise for all of the life within Canada."

Towards a better understanding of all life on Earth

CGEn's Canada BioGenome Project, funded through Genome Canada's 2020 Large Scale Applied Research Projects program and led by Drs. S. Jones and M. Murray, launched in the fall of 2021. This initiative is a natural progression from CGEn's CanSeq150 program and aims to deliver 400 reference quality, annotated genomes of Canadian species relevant to conservation planning, biodiversity, and the economy, by integrating long- and short-read sequencing approaches and through technology development. These species will be of significant value to Indigenous peoples and/or Canadians in general, and identified through a three-tiered approach: 1 – identification through ecosystem connections to three pre-selected case study species (Musk Oxen, Long-toed Salamander, Massasauga Rattlesnake), 2 – pre-selected based on importance to Canada as well as specific partners and end-users, 3 – through detailed assessment and identification based on consultation with experts, endangered status, accessibility of samples, relative impact of climate change, and contribution to the goals of the international Earth BioGenome Project. Samples will be processed and sequenced at one of CGEn's three nodes, depending on technology requirements and geographic proximity.

As the project is ramping up, several expert committees have been formed to guide the project in selecting species as well as assist with sample acquisition. Extraction (DNA and RNA) and sequencing (HiC ligation, long- and short-read) of 10 species is in progress. This includes: Pacific Spiny Dogfish, Atlantic Whitefish, Star-Tipped Cup

Lichen, Greenland Cockle, Arctic Surfclam, Northern Propeller Clam, Copper Redhorse, Arctic Char, Monterey Sea Lemon, and the Loggerhead Sea Turtle. The Pacific Spiny Dogfish is the first species with both long- and short-read sequencing data results. Further species sequencing, assembly, annotation will continue as additional species are selected.

An important aspect of this project is the intent to make the data openly available, and easily accessible to researchers and non-specialists alike. Building on cyber tools created by project team members, a prototype tool that is accessible to non-specialists has been created and is currently being tested. This tool contains terminology to enhance cross-community understanding and a geospatial database. Team members are also engaging with end-users through workshop activities that will provide the basis for building knowledge mobilization product design concepts and requirements.

The Canada BioGenome Project will also deliver policy recommendations on how to responsibly implement genomic tools in wildlife conservation, ecosystem-based co-management, species restoration, and environmental monitoring. Many of the species to be sequenced are selected based on existing and established priorities of Indigenous Peoples (e.g., the Indigenous Guardian Program), with reference to relevant policy documents of Indigenous organizations (e.g., Assembly of First Nations 2009, Inuit Tapiriit Kanatami (ITK 2018)), and with both federal and provincial end-users and conservation and wildlife groups. These organizations have a history of, or strong interest in, using genomic information to develop tools and solutions for wildlife conservation, monitoring and management. A workshop-format process is being developed by which end-users can be engaged in identifying how genomes and genomics can be used to meet their needs as well as communicating concerns and obstacles to realizing this goal.

Loggerhead Sea Turtle

Other major projects supported by CGEn (2021-2022)

Terry Fox Marathon of Hope Cancer Centres Network – ISED	Coronavirus variants rapid response network (CoVaRR-Net) – CIHR	Province of Ontario Neurodevelopmental Disorder Network (POND) – Ontario Brain Institute
Deciphering the genome biology of relapsed lymphoid cancers to improve patient management – Genome Canada	Silent Genomes: Reducing health care disparities and improving diagnostic success for children with genetic diseases from Indigenous populations – Genome Canada	MSSNG: Autism genome sequencing – Autism Speaks USA and Canada
Personalized Oncogenomics (POG) – BC Cancer Foundation	Centre for Epigenome Mapping Technologies – CIHR	Overcoming treatment failure in lymphoid cancers – Terry Fox Research Institute
The Enhanced Pancreatic Cancer Profiling for Individualized Care project (EPPIC) – Terry Fox Research Institute	Standardized and Genome-wide clinical interpretation of complex genotypes for cancer precision medicine – NIH	Spit for (Ontario) Science (Centre) – CIHR
Dissecting the effects of genomic variants on neurobehavioral dimensions in CNVs enriched for neuropsychiatric disorders – NIH	Identifying the causes of small vessel disease and stroke – ANR(France), ERC(EU), McGill	Canadian Epigenetics, Environment and Health Research Consortium Network Phase II – CIHR

"CGEn consolidates Canadian scientific leadership providing broad, critical support to the Canadian research community. CGEn cultivates the next generation of Canadian genome scientists through advanced training, ensuring that Canada is poised to respond to future large-scale genomics challenges."

Mark Lathrop, Scientific Director, CGEn-Montreal node, Professor, Department of Human Genetics and Scientific Director, McGill Genome Centre, McGill University

Looking Ahead

CGEn has become the vanguard for genome sequencing and informatics in Canadian science, supporting over 3,000 users across Canada, performing research related to human health, agriculture, forestry, the environment and more. Thanks to advances in DNA sequencing technology, relative reductions in cost, and increased demand from the research community, the amount of genome sequence data being generated by CGEn is growing at an astonishing pace (from 0.8 Pb in 2015 to over 8.0 Pb in 2022).

As we look ahead, CGEn will continue on this trajectory, supporting Canada's largest, highest profile sequencing projects for disease-associated genomes, as well as many smaller projects across the country. CGEn will continue to provide critical support for Canadian researchers by enabling genome-wide experiments at the highest resolution and accuracy, to cast the widest net for data that may generate more informed hypotheses or lead to a discovery. With CGEn's emphasis on biodiversity and conservation genomics – as exemplified by the Canada BioGenome project – we will continue to help understand the status of critical Canadian flora and fauna, including those in northern and traditional Indigenous lands.

With continued funding from the Canadian Foundation for Innovation (CFI) and other partners, CGEn will expand its role as a national facility, leading Canada and the world in research by providing the most up-to-date research infrastructure, the most advanced analytical modalities, and the best-trained personnel for genome research to thrive in Canada. Our vision is for genomics to have a positive influence on the life of every Canadian.

"We remain committed to our vision and core values of providing essential infrastructure and expertise to enhance, accelerate and support genomics research in Canada. Looking ahead, we will continue to work closely with our partners and stakeholders to forge a path towards a healthier and more sustainable future for all Canadians."

Naveed Aziz, Chief Executive Officer, CGEn

About CGEn

CGEn is a federally funded national platform for genome sequencing and analysis. Established in 2014, CGEn employs over 200 staff, and is funded primarily by the Canada Foundation for Innovation (CFI) through its Major Science Initiatives Fund (MSI), leveraging investments from Genome Canada and other stakeholders. 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), providing genomic services, including genome sequencing and analysis, that enable research in agriculture, forestry, fishery, the environment, health sciences, and many other disciplines of interest to Canadians.

Board of Directors

Gordon McCauley (Chair)
President & CEO, adMare BioInnovations, Vancouver

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

Dame Kay Davies
Professor of Genetics, MDUK Oxford Neuromuscular Centre Department of Physiology, Anatomy and Genetics, Oxford

Philippe Gros
Professor, Department of Biochemistry, McGill University, Montreal

Paul D.N. Hebert
Director, Centre for Biodiversity Genomics, University of Guelph

Bartha Maria Knoppers
Professor, Canada Research Chair in Law and Medicine & Director, Centre of Genomics and Policy, Faculty of Medicine, McGill University

Roderick R McInnes
Alva Chair in Human Genetics, Professor of Human Genetics & Biochemistry, Lady Davis Institute, Jewish General Hospital, McGill University

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

Scientific Advisory Board

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

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

Elaine R Mardis
Co-Director, Institute for Genomic Medicine, Nationwide Children's Hospital

Gabor Marth
Professor of Human Genetics, University of Utah

Roderick R McInnes
Alva Chair in Human Genetics, Professor of Human Genetics & Biochemistry,
Lady Davis Institute, Jewish General Hospital, McGill University

Richard M Myers
President and Science Director of the HudsonAlpha Institute for Biotechnology

Executive Committee

Naveed Aziz
Chief Executive Officer, CGEn

Steve Jones
Scientific Director CGEn-Vancouver, BC Cancer and University of British Columbia

Mark Lathrop
Scientific Director CGEn-Montreal, McGill University

Stephen Scherer
Scientific Director CGEn-Toronto, SickKids and University of Toronto

communications@cgen.ca
www.cgen.ca
@CGEnSeq
ca.linkedin.com/company/cgen