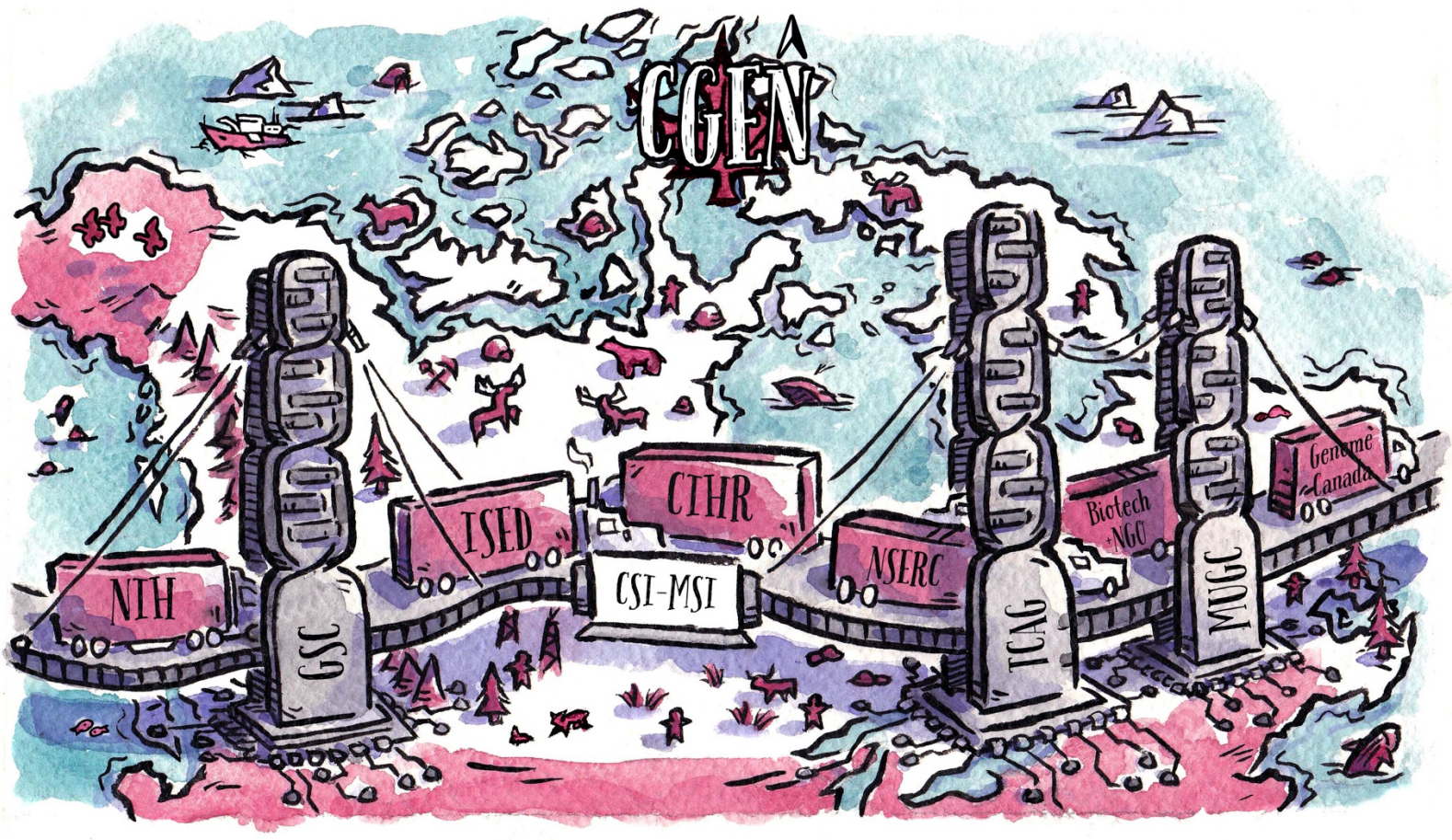

CGEn
Strategic
Vision
2025

CGEn

Canada's national platform for genome sequencing and analysis

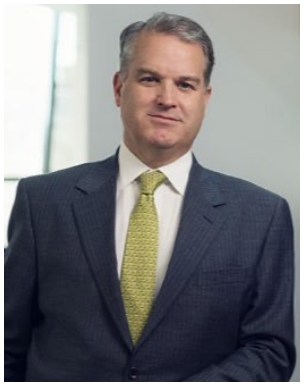


Board Chair's Message

The Board of Directors and Executive Committee of CGEn are pleased to endorse CGEn's *Strategic Vision 2025*. Since its inception in 2015, CGEn has amplified Canada's national capacity for genome sequencing and informatics analysis. *Strategic Vision 2025* builds upon a strong foundation, and outlines the key opportunities and challenges that lie ahead in genomics. CGEn is at the forefront of genomics-enabled research and discovery in Canada.

The six strategic priorities outlined in this five-year plan were formulated through strong collaborations with world-leading scientists and key stakeholders in Canada's biomedical research community. These priority areas emphasize CGEn's national imperative to support domestic biomedical research and innovation; inspire industrial transformation; advance technological development; and educate and train the next generation of Canadian technologists and scientists.

On behalf of the Board of Directors and Executive Committee, I wish to thank all who have contributed to CGEn's *Strategic Vision 2025* and look forward to working together to see its full realization.



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

CEO's Message

CGEn provides the leading genomic infrastructure for Canada's research and innovation communities. Today, thousands of scientists and researchers across Canada are capitalizing on the opportunities presented by our cutting-edge technology and large-scale data-generation platforms. Since its inception, CGEn has facilitated the sequencing of thousands of genomes in humans and other species, catalyzing groundbreaking new research in Canada.

Our *Strategic Vision 2025* represents the next phase of CGEn's leading role in genome sequencing and analysis in Canada. Growing and enhancing CGEn over the next five years will help to solidify Canada's position as a world-leader in genomics research, and help to uncover the genetic determinants of health and disease for Canadians.

Thank you to our advisory and governance boards for contributing their scientific expertise and thought leadership to CGEn's *Strategic Vision 2025*. I look forward to working with you all to promote this era of genome science research and position Canada at the forefront of genomics research in a long-term, sustainable manner.

Naveed Aziz
Chief Executive Officer, CGEn

Table of Contents

Executive Summary

CGEn is an integrated national platform for genomic sequencing and analysis. Established in 2015, the platform has since secured over $120M in public funds with a mandate to make cost-effective and high-quality sequencing (including genome sequencing) a reality for all Canadian researchers. Since its inception, CGEn has served over 2,900 user labs from across Canada and around the world and CGEn staff have authored over 565 peer-reviewed publications. CGEn's intensive work with Canadian academics, industry and government has substantiated the power of genomics in enabling discovery and commercialization at an unprecedented scale. With this solid footing, CGEn is poised to take on its next cycle of work, continuing to provide the leading genomic infrastructure for Canada's research and innovation community, but also maturing into its role as the steward of the nation's genomics enterprise on the global stage.

CGEn's *Strategic Vision 2025* was developed through a careful and objective examination of the current and future global science and tech landscape, focusing on major opportunity areas where CGEn can and should help to boost Canada's presence and impact. These priority areas, elucidated through the engagement of global thought leaders and key Canadian stakeholders, emphasize CGEn's national imperative to support domestic biomedical research and innovation; inspire industrial transformation; advance technological development; and educate and train the next generation of Canadian technologists and scientists.

CGEn's ability to harness and deliver on these opportunities is driven by four key factors. First, CGEn brings access to unprecedented domestic capacity for sequencing, allowing for the conduct of large-scale never-before-possible programs within Canada's borders. Second, governance structure that is embedded in fiduciary integrity making sure CGEn is achieving its mission goals, third, CGEn is led by globally recognized scientific leaders embedded in the academic structure of three world-class institutions (SickKids Hospital, McGill University and BC Cancer Agency), affording the platform instant credibility and cutting-edge capabilities. Fourth, CGEn's critical mass of expertise and infrastructure is helping to attract, integrate and synergize the multitude of solitary genomics efforts across the country into one cohesive and globally competitive force. Last, the standardization of systems across CGEn's multiple distributed nodes allows Canada to circumvent critical quality and integrity issues that have plagued other leading jurisdictions in the context of the global genomics race.

About CGEn

CGEn is a federally funded 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 (MSI), leveraging investments from Genome Canada and other stakeholders. CGEn operates as an integrated national platform that federates the major genome science centres 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) in a single platform. CGEn provides state-of-art genomics, including genome sequencing and analysis, that enable research in agriculture, forestry, fishery, the environment, health sciences, and many other disciplines of interest to Canadians.

CGEn's primary mission is to consolidate Canadian leadership in genomics by dramatically expanding sequencing and computational capacity nationally, and training highly qualified personnel so that the most important genome projects to Canada's health, environment, and economy can be conducted competitively by our own researchers. CGEn is administered centrally, with infrastructure bridging the country, providing enhanced regional access for Canadian scientists, allowing for co-funding from provincial governments, and leveraging additional stakeholders including host institutions. Since its inception in 2015, CGEn has facilitated the sequencing of thousands of genomes, in humans and other species, catalyzing groundbreaking new research.

Dedicated to "open access" science, in 2020, CGEn supported over 1,953 Principal Investigator laboratories, from all 10 Canadian provinces. These include essentially all research-focused Canadian universities, many colleges, and over 100 provincial or federal government organizations and NGOs. CGEn provides services to projects funded by Canadian funding agencies like Genome Canada, the Canadian Institutes of Health Research (CIHR), the Natural Sciences and Engineering Research Council (NSERC), international funding sources such as the U.S. National Institutes of Health (NIH), the EU Horizon 2020, the Wellcome Trust and other national funding agencies, provincial governments, disease charities, and biotech companies like Deep

Genomics, Ambry Genetics and Bionano Genomics use CGEn infrastructure. In addition, in 2020 CGEn staff authored over 130 peer-reviewed publications, was cited as a contributor in approximately 300 peer-reviewed publications, and helped to train hundreds of undergraduates, Masters and PhD students, Post-doctoral Fellows and technical and professional staff.

The three CGEn nodes operate as an integrated world-leading facility, with top-of-the-line instrumentation with a combined sequencing capacity of over 40,000 (30X) human genome equivalents per year. The CGEn platform is supported by equally robust computational capabilities, including more than 50,000 compute cores and over 50 petabytes of storage. CGEn develops and utilizes state-of-the-art bioinformatics tools and software to enable all stages of genomic data processing, from the primary analysis of raw sequences to tertiary derivation of application-relevant insights. One of CGEn's unique values rests in a standardized model of distributed delivery of high-quality genomics adapted to the Canadian research environment. CGEn allows researchers from across Canada or nation-wide programs, to access the highest-quality sequencing and analysis pipelines with secure data outputs from any one, or a combination of, its nodes. The creation of CGEn has already catalyzed important new national projects and secured Canadian leadership in major international initiatives in genomics. With this solid foundation, CGEn is poised to continue to push the boundaries of genomic science, and to advance in its role as Canada's genomics powerhouse and the steward of the nation's genomics data enterprise. Over the next five years, CGEn's work, commitments and deployment of resources will centre around six opportunity areas (see below, Strategic Priorities).

Vision and Mission

VISION

- Serve as Canada's engine for genomics-enabled research and discovery.

MISSION

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

The Upcoming 5 Years

CGEn's intensive work with academics, industry and government has substantiated the power of genomics in enabling discovery and commercialization at an unprecedented scale in Canada. With this solid footing, CGEn is poised to take on its next cycle of work, continuing to provide the leading genomic infrastructure for Canada's research and innovation community, but also maturing into its role as a steward of the nation's genomics enterprise on the global stage.

CGEn's Strategic Vision for the next five years was developed through a careful and objective examination of the current and future global science and technology landscape, focusing on major opportunity areas where CGEn can and should help to boost Canada's presence and impact. These priority areas emphasize CGEn's national imperative to support Canadian research and innovation, inspire industrial transformation, advance technological development, and educate and train the next generation of Canadian technologists and scientists.

CGEn's ability to harness and deliver on these opportunities is driven by four key factors. First, CGEn brings access to unprecedented domestic capacity for sequencing, allowing for the conduct of large-scale, never-before-possible programs within Canada's borders. Second, CGEn is led by global scientific leaders embedded in the academic structure of three world-class institutions (The Hospital for Sick Children, McGill University and BC Cancer), affording the platform instant credibility, providing access to resources and collaborations that leverage institutional capabilities. Third, CGEn's critical mass of expertise and infrastructure is helping to attract, integrate and synergize the multitude of solitary genomics efforts across the country into one cohesive and globally competitive force. Last, the standardization of systems across CGEn allows Canada to circumvent critical quality and integrity issues that have plagued other countries where Genome Centres are competing against each other, while retaining a presence across the country and important regional and provincial linkages.

Over the next five years, CGEn will continue to increase its sequencing and informatics/data storage capacity, working in conjunction with federal, provincial and private-sector partners. CGEn will work with relevant stakeholders, host institutions, and partners in order to initiate and support large-scale, genome sequencing cohort studies. To secure Canada's role as global leader, CGEn is taking leadership in the conception and planning of large Canadian genome sequencing studies and is building upon lessons learned from successful initiatives such as the 10,000+ genome MSSNG autism project, while also implementing new approaches and technologies developed within CGEn. Hand in hand with these capacity increases, CGEn will enable greater delivery of data to end-user scientists, more sophisticated consultation with clients on project design and data analysis and delivery. CGEn will promote and enable the general philosophy that every biological investigation should be accompanied with a genome sequence. CGEn will also continue to monitor advances in the sequencing field, evaluate new technologies as they arise, and implement those of most benefit to Canadian science. CGEn will continue to expand its network of users and collaborators by leveraging its internal scientific expertise and input from world-leading scientists who have agreed to serve on its governance and advisory boards. CGEn will interact with other major national initiatives that are supported by CFI or other funders to enable integration of scientific research in our country (such as with the Canadian Longitudinal Study of Aging as described below).

This strategic plan outlines the major goals and challenges for CGEn over the next five years, as we see them now. However, to remain competitive, it is imperative that CGEn remain nimble, and adjust its goals both proactively and reactively as this very quick-moving field evolves.

STRATEGIC priorities

1. Participate in large-scale human genomic data generation programs and catalyze the development of a national cohort study that would seek to characterize the vast diversity of Canada's population.

2. Create and assemble clinically relevant data, algorithms, technologies and methodologies in medically relevant areas, such as rare diseases, paediatric cancer, autism and other neurodevelopmental disorders, health and disease in aging, bringing shared ideas and datasets to enhance applications in medicine across Canada.

3. Develop methodologies for sequencing non-human species that are of critical relevance to the ecology, economy, culture, health and wealth (collectively, biodiversity) of Canada.

4. Develop powerful and efficient solutions for big data storage, data deposition to "open-science" databases, and high-speed data exchange and genome sequence analysis.

5. Lead technology development through innovation, validation and integration of new genomic methodologies and establish best practices in key application areas.

6. Educate, diversify, and enrich the Canadian genomics talent pool to continue to deliver on its ambitious milestones, and to support the growing need for these researchers in clinical and biological sciences.

GOALS & STRATEGIES 2021-2025

01 **DATA GENERATION:** In today's research landscape and increasingly into the future, high-value biomedical discovery is dependent on large-scale genomic data generation.

In order to fully harness the opportunities that are presented by genomics in the coming years, CGEn is identifying domestic and international opportunities for large-scale data generation with its platforms. We have already made substantial headway in this regard, leading and participating in numerous big-data projects that seek to unpack the genetic determinants of health and disease for Canadians. We illustrate these efforts through a number of examples.

CGEn is engaged with The Canadian Longitudinal Study on Aging (CLSA) cohort, which is our country's flagship population cohort study, funded by the Canadian Institutes of Health Research (CIHR) and CFI. In partnership with Public Health Agency of Canada, Health Canada, Economic Social Development Canada and Veterans Affairs, CLSA is following 50,000 people between the ages of 45 and 85, from all provinces, with committed public funding for at least 20 years to obtain follow-up clinical exams and new biological samples collected on a regular basis. The rich information being generated includes comprehensive genome-wide genotyping and sequencing data, epigenomic, metabolomic and proteomic profiling at multiple time points as the cohort ages. Genotyping and sequence data will provide information on stable germ-line genetic variation and its relationship to epigenomic patterns, metabolites and proteins, and also somatic variation (such as clonal haematopoiesis and telomere length) that may have important relationships with aging and disease outcomes of aging.

CGEn is also working closely with other established population health programs in Canada to enrich their data pools with valuable genomic information. For instance, in a recent project in collaboration with the Ontario Brain Institute (OBI), genomic data from 97 children with a common form of cerebral palsy were analyzed against 10,000 controls to hone in on a set of rare genetic mutations involved in brain development and function, and in the Quebec-based

GENetics of Glycemic regulation in Gestation and Growth (Gen3G) birth cohort study we are using genome sequencing to investigate the pathophysiology of impaired glycemic regulation in pregnancy including genetic determinants, and to evaluate the impact of these factors on metabolic outcomes in offspring.

As another example, in the early years of its formation, CGEn facilitated the mid-point (2015) repatriation of the MSSNG project, which aimed to elucidate the genetic diversity of Autism among a cohort of 10,000 children and their families, many being Canadian. After 3,000 MSSNG samples were sequenced abroad (due to lack of capacity and appropriate infrastructure in Canada), only one year after its formation, using its distributed capacity, CGEn took on sequencing and analysis of the remaining 7,000 specimens, already helping to uncover 61 genetic and chromosomal loci related to the pathogenesis of this complex neurodevelopmental disorder which affects one in 66 Canadians. Through initiatives like the MSSNG project, CGEn has also led the creation of an open-access genomic data portal, which is powered by cross-sectorial collaborations between the non-profit (Autism Speaks), private (Google, Verily and Canadian informatics company DNAStack), tying together 17 leading clinical researchers and more than 3,500 families affected by autism, across the entirety of Canada.

Additionally, CGEn has sparked and leads the Personal Genome Project Canada (PGP-C), to give a select group of participating Canadians a fully 'open-science' lens into their genetic make-up and how their genomes may impact their risk of developing disease. The program is deep in concept, but smaller in scale, already having broad implications for healthcare delivery in Canada. In the first highly newsworthy analysis of PGP-C data, it was found that 25% of the 56 participants harboured genetic variants with actionable health indications, and over 10% harboured pathogenic or likely pathogenic variants.

To facilitate data generation from diverse groups, CGEn is participating in DNA sequence studies that seek to ensure equitable treatment and study of indigenous populations across Canada. For example, the Silent Genome Program sets out to catalog genetic diversity amongst indigenous populations to ensure the similar power to detect causative genes of rare diseases and provide accurate genetic diagnoses as in other ancestral backgrounds.

CGEn sites have been major contributors to the International Human Epigenome Consortium (IHEC) through the CIHR/CEEHRC Epigenetics, Environment and Health Signature Program. We have used this positioning to organize large national and international consortia targeting genomic and epigenomic studies of chronic diseases with funding from international sources including pharma. An example is the BRIDGET project, a Joint Program for Neurodegenerative Disease consortium to identify biomarkers of cognitive decline with funding from several countries, supports both research and core platform activities at CGEn for genome sequencing and methylation profiling of many thousands of samples from cohorts with neuro-imaging data. Similarly, the "Oncohistone" project, funded by NIH/NCI to undercover histone variation and modifications in paediatric brain cancer, is supported by CGEn. Members of CGEn teams lead international consortia to determine the role of epigenome changes in metabolic disease ("EpiTRIO") and chronic inflammatory disease ("Reset-

AID), each with multimillions in funding from Australia, Canada, Germany and France. Through these and other initiatives, we have observed a rapid increase in epigenome and other sequencing activity related to functional genomics over recent years. To our knowledge, CGEn is now the largest producer worldwide of sequence-based methylomes for human population and epidemiological studies.

To accelerate progress in microbiome research, as part of its own research programs and through collaborative research, as well as through focused technology development, CGEn will develop improved molecular and computational methodologies for comparison of microbiome and human genomes with human health phenotypes. Through its deep network with other Canadian institutions, CGEn will make these advances available to the scientific community. This will directly accelerate microbiome research in Canada by ensuring standardized methodologies of the highest quality for all researchers. In doing so, CGEn will catalyze microbiome research in Canada and help to ensure that the nation benefits from the microbiome revolution. It will be through such large-scale collaborative efforts that CGEn can help to advance the impact of microbiome in personalized medicine in Canada and internationally.

Through our support of research such as described above, CGEn is making a major contribution to the creation of a national "big data platform" that will link genome sequence with other large-scale biological data, such as metabolomics and proteomics, with environmental/lifestyle information to address questions of major public health significance: for example, to discover biological and/or environmental markers that will predict diseases such as cancer, diabetes, and dementia before they appear, and change outcomes, including finding predictors of individual responses or adverse reactions to medication. These data will allow for the dissection of common diseases, which are made up of heterogeneous disease sub-types, each requiring their own diagnostic and treatment strategies. A major application will be to understand how, individually and in combination, genetic variation and environmental exposures impact health maintenance and the development of disease and disability as people age, and to use this information to inform health and social care policies in Canada. Canadian research in rare disease, paediatric conditions, and mental health, across ages, will also immediately benefit. The CGEn sites encompass major Canadian efforts on the ethical, legal and policy issues associated with this research, and our investigators have leadership roles within many associated international initiatives such as the Global Alliance for Genomics and Health (GA4GH), the Public Population Project in Genomics and Society (P3G), the International Human Epigenome Consortium (IHEC), and the Human Cell Atlas (HCA). We are using this expertise to provide a framework for genomic data generation and its distribution within Canada.

Building on its existing work in this area and over the coming five years, CGEn will continue to participate in large-scale human genomic data-generation programs and catalyze the development of a national cohort study that would in part seek to characterize Canada's Indigenous and founder populations.

02 **CLINICAL GENOMICS:** Whole genome sequencing is being actively integrated into clinical decision-making to improve patient outcomes.

While Canada has not yet established a national directive for embarking on genomics-based healthcare delivery, CGEn's research work in this area is laying a solid foundation. For instance, the PROFYLE project, a pan-Canadian research effort between 30 paediatric research and funding organizations (which is in part led and powered by CGEn), is using genome sequencing to identify molecular biomarkers of childhood cancers. PROFYLE aims to enroll 450 patients in the first four years of its work, to sequence and characterize their tumours with support from a committee of experts, and to help guide these subjects' care toward better outcomes. PROFYLE's clinical recommendations include more targeted forms of treatment and enrollment in appropriate clinical trials.

CGEn's work has helped uncover many other opportunities where genomic information can help to change the course of care and transform patient outcomes. The Personalized Oncogenomics Program (POG) in Vancouver has sequenced the genomes and transcriptomes of tumours from almost 1,000 patients with the goal of providing insights into the cancer driving pathways and relevant therapeutic approaches. The CGEn-powered Kids Cancer Sequencing Program (KiCS) at SickKids Hospital in Toronto is another key example. The program has enrolled 70 patients with new or relapsed cancer, and with the help of genome sequencing, determined that a subset of childhood bone cancers are detectable much earlier than they are currently diagnosed. At McGill University in Montreal, CGEn is supporting the Genome Canada Large Scale Applied Research Project (LSARP) titled "Tackling childhood brain cancer at the root to improve survival and quality of life" which is co-led by Drs. N. Jabado at McGill, M. Taylor at SickKids and J. Majewski at McGill, focussing on discovering genomic vulnerabilities of the most aggressive childhood brain tumours and translating those discoveries into personalized therapies. Also, CGEn plays a key role in driving the McGill Clinical Genomics Initiative forward, where a pilot phase to perform genome sequencing on every new breast cancer patient at the McGill University Hospital Centre is now underway (900 patients). The aim of this initiative is to perform genome sequencing on every patient with cancer, cardiovascular, metabolic or rare disease in the next five years.

Additionally, all three of CGEn's nodes make scientific and sequencing contributions to the Care4Rare program—

a pan-Canadian multidisciplinary team of over 300 physicians and scientists working on the application of genomic technology in rare disease. Care4Rare aims to identify the unknown genetic causes of rare diseases through the integration of genomic sequencing into routine diagnostic workflows. Thus far the program has helped diagnose over 1,000 different rare diseases (including 82 never-before-described disorders) in 5,000+ patients from around the world[1].

CGEn's capacity and leadership in the clinical genomics space has been critically bolstered by the work and reputation of its institutional homes: The Hospital for Sick Children, McGill University and BC Cancer. Moreover, CGEn's establishment as a standardized, cohesive cross-national platform has created a nucleus for coordination and synergistic amplification of numerous other initiatives within the boundaries of its home institutions that are tackling various aspects of clinical genomics, from research to ethics to delivery.

Over the coming five years, CGEn will help to standardize, leverage and bring access to clinically relevant data, algorithms, policies, technologies and methodologies in select disease areas such as paediatric cancer and rare diseases. In so doing, CGEn will bring ongoing initiatives together into a cohesive effort for the implementation of genomic medicine in Canada.

[1] https://www.ncbi.nlm.nih.gov/pubmed/28170084

03 **BEYOND HUMAN:** Genomic sequencing of non-human species is becoming an increasingly pertinent tool for driving industrial transformation in the 21st century.

Like other countries, Canada is recognizing that to remain a player on the global stage, its core economies (namely natural resources and agriculture) must undergo at least some degree of re-invention. CGEn has been working intensively over the last few years to position genomics as a key enabler of this impending revolution. For example, in an effort to support biodiversity, conservation, agriculture and breeding, CGEn has recently launched the CanSeq150 initiative[2] —a nomination-based program to sequence genomes of 150 species relevant to Canadian culture, science, and industry. CanSeq150 is already about two-thirds of the way to reaching its 150 species targets. Building from the foundational sequencing of the genome of Canada's national animal, Castor canadensis (the Canadian beaver)[3], examples of other species being sequenced include the artic fox, northern bottlenose whale, northern fur seal, grizzly bear[4], Canadian wood bison, lake sturgeon, Vancouver Island marmot, black spruce, the Greenland flounder and the beluga whale. This initiative has captured the attention of hundreds of biologists across all major provinces of Canada including many universities, Federal and Provincial departments like Fisheries and Oceans Canada and Natural Resources Canada, The Toronto Zoo, The Royal Ontario Museum, The Vancouver Aquarium and Assiniboine Park Zoo. Through the CanSeq150 initiative, CGEn is also working with the International Barcode of Life (iBOL)[5] and the Earth BioGenome programs to help discover and sequence species that are important in the Canadian context.

CGEn has also fueled numerous large-scale programs to help optimize livestock, agriculture and aquaculture production in Canada. CGEn scientists have also contributed heavily to the sequencing and assembly of the poplar tree genome, to help conservationists optimize breeding strategies, and ultimately improve Canadian forest health and wood quality. For example, generating the first genome of the white spruce, which represented at the time one of the largest genomes ever assembled and one of the first conifer genomes[6][7]. Conifer genomes being sequenced currently by CGEn also include those of the Sitka and Engelmann spruce.

Over the coming five years, CGEn will continue to advance projects and hone methodologies and analytical pipelines for sequencing non-human species that are of critical relevance to the environment, health, and wealth of Canada.

[2] http://www.cgen.ca/canseq150-overview
[3] https://www.g3journal.org/content/7/2/755
[4] https://www.ncbi.nlm.nih.gov/pubmed/?term=30513700
[5] http://ibol.org/
[6] https://www.ncbi.nlm.nih.gov/pubmed/?term=23698863
[7] https://www.ncbi.nlm.nih.gov/pubmed/?term=26017574

04 **COMPUTATIONAL GENOMICS:** Influx of large amounts of genomics data underscores the need for new computational tools to aid the analysis and integration of complex data sets to better understand the associated biology.

High-throughput technologies, and in particular next-generation sequencing (NGS), continue to revolutionize biomedical research by enabling the genetic and genomic characterization of the molecular processes on both organism and cell level with unprecedented resolution. Although these developments promise to have a significant impact on life science research and health care, the immediate challenge of storing, processing, analyzing and sharing the vast volumes of data generated by these platforms represents a major bottleneck. These issues require appropriate planning at the level of compute infrastructure, data management and research software.

CGEn and its partners have developed a rich expertise in computational genomics, which will be harnessed to address these challenges. For example, Dr. Michael Brudno (The Hospital for Sick Children, Toronto) has been developing popular software tools to capture and analyze phenotyping data[8] and Dr. Steven Jones (Canada's Michael Smith Genome Sciences Centre, Vancouver) has been developing advanced tools for the study of cancer datasets such as CancerMine[9]. Similarly, Dr. Guillaume Bourque (McGill Genome Centre, Montreal) has been developing a suite of scalable pipelines for genomic analyses[10].

Beyond software to analyze genomic data, a major trend spearheaded by the Global Alliance for Genomics and Health (GA4GH) is to pursue the development of strategies to enable large-scale sharing of genomic data in an ethical and secure way. With projects led by CGEn Principal Investigators, such as the Canadian Digital Infrastructure for Genomics (CanDIG, https://candig.github.io/), we are in a unique position to advance ways genomics data can be shared and analyzed not only nationally but also internationally.

Over the coming five years, CGEn will continue to establish powerful and efficient solutions for big data storage and high-speed data exchange and standardize tools and workflows to serve diverse genomics applications across the country.

[8] https://www.ncbi.nlm.nih.gov/pubmed/23636887
[9] https://www.ncbi.nlm.nih.gov/pubmed/31110280
[10] https://academic.oup.com/gigascience/article/8/6/giz037/5513895

05 **TECHNOLOGY PUSH:** The boundaries of genomics science are being constantly pushed by the emergence of novel sequencing technologies and analytical tools.

In Canada, CGEn is leading the development of novel methodologies (from sample processing to data mining) to help advance the frontiers of genomics science with the latest available sequencing technologies. For example, CGEn has produced some of the earliest literature on key topics such as RNA sequencing and histone modifications. More recently, CGEn nodes have played a central role in implementing advanced epigenomics protocols and generating data for a CIHR strategic initiative called the Canadian Epigenetics, Environment and Health Research Consortium. In this context, we have also developed a data portal which now supports the International Human Epigenome Consortium[11].

When it comes to second generation sequencing technologies, currently Illumina's NovaSeq6000 is the most powerful high-throughput production scale sequencing system to date. It has allowed the timely completion of existing genomic sequencing projects and has facilitated the uptake of novel sequencing-based applications. In the next two years, CGEn will seek funding to ensure the full transition from the previous IIlumina sequencing systems (HiSeqX) to the newer NovaSeq system. Over the next five years, CGEn will actively evaluate other upcoming sequencing technologies, including Illumina's recently introduced new semiconductor-based signal detection technology (iSeq 100) and MGI's production scale instruments (MGISEQ-T7) offering novel detection technologies that have recently entered the market.

Also, of considerable interest to CGEn are the latest third generation sequencing technologies capable of sequencing single DNA molecules in long reads, with commercial systems available by PacBio and Oxford Nanopore. New technologies will provide insights into sequence information across complex regions of the genome and characterize complex sources of variation while vastly increasing the speed of DNA sequencing and lowering the cost. Since the utility of these methods relies on the ability to obtain high molecular weight (HMW) DNA from diverse sample sources, CGEn is also focusing on optimizing DNA isolation techniques on magnetic disks and microfluidic chips using a combination of fractionation instruments (Circulomics Nanobind, SAGE HLS; SAGE Blue Pippin) and sharp cut-off points to obtain high quality very HMW DNA molecules. To improve the intactness of the DNA, CGEn is developing "on-chip" NGS library preparation approaches that will allow us to perform DNA isolation and library prep in one go on the SAGE HLS microfluidics chip. CGEn nodes are also exploring similar microfluidics techniques for CRISPR-based approaches that will allow for streamlined isolation of loci of interest from HMW DNA and sequencing library prep, thus bypassing hybridization-based approaches.

[11] https://www.ncbi.nlm.nih.gov/pubmed/27863956

CGEn is also evaluating fourth generation single cell sequencing approaches. The scRNA-seq and scDNA-seq methods established at CGEn rely on separation of single cells from tissue by enzymatic or mechanical dissociation, resulting in loss of spatial information. Very recently, fourth generation sequencing technologies that preserve the spatial coordinates of RNA and DNA with subcellular resolution have emerged. The technologies are currently based on short read sequencing and are able to map back sequencing reads to their original location in a given tissue using barcode technology. These methods include approaches developed by the Lundeberg lab and marketed by the company "Spatial Transcriptomics," now commercialized by 10x Genomics, while Nanostring has developed a competitor approach "Digital Spacial Profiling" using a new instrument, The GeoMx. CGEn will aim to establish at least one of these technologies along with single cell genomics. Other technologies including methods based on padlock probes combined with rolling circle amplification or FISSEQ, which have been published over the last five years, are alternatives that will be examined as appropriate.

CGEn will continue to lead technology development in the field of genomics, working to validate emerging sequencing platforms and disseminate best practices for sample processing.

06 **EXPANDING ROLE OF GENOME CENTRES:** Outreach, education and training are no longer auxiliary capabilities but essential responsibilities and key mechanisms of sustainability for genome centres around the world.

The genome centre model, which stipulates the centralization and delivery of cutting-edge sequencing infrastructure and analytics expertise, is one that has been adopted worldwide. Part and parcel of this model is a responsibility to not only deliver high-quality and timely service, but also to provide scientific leadership, support for experimental design, and to establish and disseminate best practices and standards, and communicate the latest advancements in the field. These activities also help to seed and train the next generation of scientists and technologists who are key to the success of genome science at large.

As Canada's integrated platform for genomic sequencing, CGEn is committed to the training and education of young scientists and its users, not only to maintain the state of knowledge and standards of practice within the Canadian genomics community, but also to cultivate and foster its internal talent pool. Staying true to CGEn's equity, diversity and inclusion (EDI) principles[12], CGEn will work to ensure that its ecosystem reflects the rich diversity of the communities it serves, and that diverse voices are elevated and equitably integrated into all of its work. CGEn recognizes that a breadth of perspectives, skills and experiences contributes to excellence in research.

CGEn also provides several opportunities for training graduate students within and outside of standard M.Sc. and Ph.D. programs in human genetics, as exemplified by the postgraduate Laboratory Course in Genomics at McGill where CGEn-funded personnel were directly involved in student supervision. At its Vancouver node, CGEn delivers a graduate-level Genome Science + Technology course that combines foundational principles of genomics with the latest technological advancements in the field. At CGEn-Montreal node, staff teaches genomic technologies in undergraduate courses in Medical Genetics as well as Bioengineering. CGEn also participates in the Queen Elizabeth II Diamond Jubilee Scholarships program—providing training at the intersection of genomics and quantitative biology—and the Kyoto-McGill Top Global University Project, a collaborative training program focused on computational methods for big-data mining in medicine and biology. CGEn also collaborates with other relevant educational initiatives in Canada, such as the Canadian Bioinformatics Workshop Series, which provides in-depth programming in computational biology as applied to specific areas of research. Further international expansion of these programs will be sought in the near future in collaboration with agencies in the UK and Germany.

Over the next five years, CGEn will work to educate and train users, enrich the Canadian genomics talent pool and cultivate internal knowledge and personnel through the fostering of user groups, vendor programming and partnerships with academic educational programs.

[12] http://www.cgen.ca/edi

CONCLUSION

CGEn, established in 2015, is Canada's national platform for genome sequencing and analysis, encompassing the three major genome centres in Toronto (The Centre for Applied Genomics/TCAG at The Hospital for Sick Children), Montreal (McGill Genome Centre/MGC) and Vancouver (Michael Smith Genome Sciences Centre/GSC).

CGEn's foundational vision is that *every biological research project will be enlightened by knowledge of an underlying genome sequence*, and that ultimately, every Canadian's genome will eventually be an integral part of her or his medical record. CGEn's operations and impact are structured to prepare for analysis and interpretation of the nation's genomic data, with rapid advances towards realization of this vision.

CGEn *"bridges"* the nation from coast-to-coast-to-coast, providing cutting-edge, cost-effective infrastructure for genomic data generation and bioinformatic analyses, ensuring major genome projects can be conducted by Canadian scientists within our country. CGEn is supported by CFI through its Major Science Initiatives Fund (CFI-MSI; 2017-2023), and leverages investments from Genome Canada and other stakeholders. With funding from CFI's Innovation Fund (CFI-8), CGEn established a sequencing platform with a capacity for tens of thousands of human genome equivalents per year. The CGEn platform is supported by robust computational capabilities, including more than 40,000 compute cores and over 50 Petabytes of storage. CGEn nodes are also active nodes of Compute Canada, further enabling information exchange and analysis of massive genome sequence datasets. CanDIG, a national architecture for distributed analysis of private genomic data, is fully integrated across the CGEn *"piliers/pillars".*

Dedicated to "open-access" science and continuing to support burgeoning demand for genomics research, in 2020 CGEn served researchers from all 10 provinces, billing over $39 million in services. Current users include essentially all research-focused Canadian universities, many colleges, and more than 100 provincial or federal government organizations such as Agriculture Canada, Fisheries and Oceans, and the Public Health Agencies of Ontario, Quebec, and Canada, as well as special projects supported by Innovation, Science and Economic Development Canada (ISED). CGEn also supports CIHR, Genome Canada and its provincial/regional centres, NSERC, international grants (e.g. NIH), disease agencies, and biotech and pharma companies.

Since its inception in 2015, CGEn has established itself as Canada's national provider of sequencing and other genomics, along with the associated informatics support, linking thousands of researchers using this platform across the country. In its next cycle of work, CGEn will continue to lead the development and integration of leading-edge technology, hone its delivery of data, and further increase its national and international reach. The strategic vision laid out in this document is CGEn's first deliverable in this role, outlining key areas of development in the field of genomics and delineating CGEn's goals and directives for helping Canada to capitalize on these growing opportunities. Our ultimate goal is to help elevate genomics research to a new level in Canada, and to catalyze the next round of innovative genome-based science.

Contact
Email: communications@cgen.ca
Twitter: @CGEnSeq
LinkedIn: www.linkedin.com/company/cgen/
Website: www.cgen.ca