

Genomics, the Digitization of Biology, and Canada's AI Opportunity

We are living through profound technological change. While breakthroughs in artificial intelligence (AI) large language models are making headlines, an equally transformative—and closely linked—revolution is unfolding in the life sciences. The digitization of biology through genomics is fundamentally altering the foundations of entire sectors, from health care to agriculture, to natural resources, to the blue economy. Unlocking the full potential of the biorevolution driven by genomics and the rise of AI requires the mutual reinforcement of these transformative general-purpose technologies.

Genomics is the science of deciphering and understanding an organism's entire genetic information encoded in DNA and related molecules. This includes humans, other animals, plants and microorganisms such as bacteria, viruses and fungi.

It took about 13 years to sequence the first human genome at a price tag of around US\$3 billion. Today, a human genome can be sequenced in about a day and costs only a few hundred dollars: a figure that continues to fall!¹ This rate of decline has far outpaced the fabled Moore's Law (doubling computing power every two years in semiconductors), underscoring just how revolutionary genomics research has been. (See Figure 1)

¹ "The Cost of Sequencing a Human Genome," National Human Genome Research Institute, 2021-11-01, <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>.

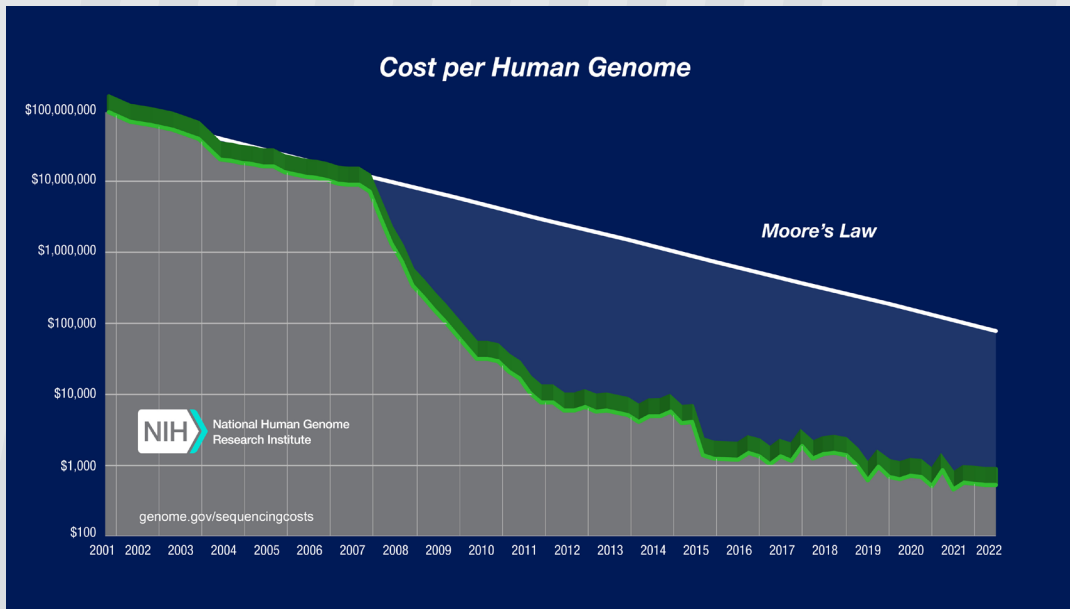


Figure 1: Source: DNA Sequencing Costs: Data: Kris A. Wetterstrand, "DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program," *National Human Genome Research Institute*, 2023-05-16, <https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data>

Unlocking the power of genomics ultimately comes down to data.

It is a highly data-intensive field. A single human genome contains around 3 billion base pairs of DNA—with associated data taking up as much as 200-300 GB of space. The rapid fall in costs has meant an exponential increase in the sheer amount of data being created: tens of exabytes (billions of GB) annually to store only the genomic data so far.²

This digitization of biology has powered the use of genomics in new research areas and the quick uptake of clinical genomics. It has led to the development of cutting-edge technologies and innovations applicable across sectors, including public health, food and natural resources. Technologies stemming from genomics, such as CRISPR and mRNA vaccines, are nothing short of transformational.

Canada's long-standing investment in genomics research across various sectors—including health, agriculture, environment, forestry, fisheries, energy and mining—since Genome Canada's founding in 2000 provides a strategic advantage over many global counterparts who have primarily concentrated their genomics efforts in the health sector.

However, the sheer scale of data production stemming from genomics poses challenges. Researchers have argued that "the rapid accumulation of genomic data from DNA sequencing and clinical data from [electronic health records] poses significant challenges and opportunities for data scientists to extract biologically or clinically relevant information from the vast amount of genotype and phenotype data."³

² "Genomic Data Science," *National Human Genome Research Institute*, 2022-04-05, <https://www.genome.gov/about-genomics/fact-sheets/Genomic-Data-Science>.

³ Jing Lin and Kee Yuan Ngiam, "How data science and AI-based technologies impact genomics." *Singapore Medical Journal* 64(1): p 59-66, 2023-01, <https://doi.org/10.4103/singaporemedj.smj-2021-438>.

The key challenge for the next phase of genomic research is turning this data into actionable insights that drive real-world outcomes in health care and beyond. Cue the convergence of genomics with AI.

AI and genomics

Given the vast amounts of data produced by researchers and clinicians, genomics is increasingly intertwined with AI, with each technology helping unlock the potential of the other.

Already, most genomics research is using AI. This includes projects where:

-
- + AI is the **objective** that **defines** the project
-
- + AI is an **activity** that **contributes** to the overall objective of the project
-
- + AI is a **tool** that is **used** within the project

Genome Canada funded its first AI-related project in 2018 through its Bioinformatics and Computational Biology Program. This program has since funded more than 75 projects, with a total investment of more than \$70 million. These projects have spanned multiple sectors and are led by researchers across the country representing all six of Canada's [regional Genome Centres](#). They also include individual investigator-driven projects and major foundational initiatives that are key to unlocking the potential of genomics through the use of computational biology, machine learning and AI. They have collectively positioned Canada as a leading global player at the intersection of these technologies.

Individual projects leveraging computational biology include the work of teams led by [Dr. Fiona Brinkman](#), [Dr. Gary Van Domselaar](#) and [Dr. William Hsiao](#), which utilized machine learning and genomics to enhance real-world health outcomes. They developed a computational platform, Integrated Rapid Infectious Disease Analysis (IRIDA), to enable more genomics-based analysis of infectious disease outbreaks. This tool is now used at Canada's National Microbiology Laboratory and for international collaborations across four continents.

Meanwhile, AI-related research also supports other areas of Canadian industrial strength, such as Canada's \$6.17 billion dairy industry. Genomics and AI are working together to deliver predictive models for dairy production, supporting management practices and optimizing dairy farm profitability, in a [Genome Canada-funded partnership](#) between Lactanet, My Intelligent Machines and research teams led by Dr. Abdoulaye Baniré (Université du Québec à Montréal) and Dr. Marc-André Sirard (Université Laval).

AI-genomics projects are also leading to commercialized technologies and spin-offs. [Creative Molecules Inc.](#) is a spin-off from [Dr. Christoph Borchers's Genome Canada-supported project](#), which developed an algorithms and software to interpret mass spectrometry data to determine 3D structural models. The work has also led to the creation of three web servers that have made protein structure tools freely available to researchers, creating wider social value from this research.

Globally, funders have been focused on activating AI and genomics' potential. Under the previous administration, the U.S. National Institutes of Health [invested US\\$130 million in the Bridge2AI program](#), which seeks to accelerate the use of AI by the biomedical and behavioural research communities. Genomics England has also launched major funding streams related to AI, including a [£26 million 'Cancer 2.0' program](#), which aims to improve the accuracy and speed of diagnosis for cancer patients through evaluating cutting-edge genomic sequencing technology—using AI to analyze genomic data alongside digital histopathology and radiology images.

Investing in Canada's sovereign data assets

Canada is also powering genomics and AI advancement beyond the level of individual research projects. Significant new initiatives are laying the data and governance frameworks to make an AI-powered genomics revolution possible—and to deliver the high-quality, diverse and dependable data assets needed to feed meaningful AI solutions as Canada ramps up the [Pan-Canadian Artificial Intelligence Strategy](#).

The availability of quality data, accessible to researchers and industry through solid governance models, is essential. The rapid digitization of biological data already “challenges existing governance models” both within countries and internationally, given that the “pace of scientific capability far, far outpaces the ability to govern.”⁴


There is growing recognition of the need to address Canada's highly fragmented data ecosystem, aligning data standards and databases to enhance interoperability. As the Government of Canada heard as part of recent consultations, the vast troves of data being produced tend to “reside in institutional and jurisdictional silos where the data standards and databases used to manage the data are not aligned and lack interoperability.”⁵ A lack of interoperable datasets holds back potential applications of AI to support genomics research.

Meanwhile, heightened geopolitical threats and global backlashes against science have also underscored the importance of ensuring access to sovereign data assets. This year, several foundational U.S. datasets, including several vital health databases, have been taken offline. Protecting Canada's data supply chain by providing access to organized data relevant to Canadian interests is a strategic imperative and essential to realizing the potential stemming from the digitization of biology.⁶

4 Smyth, Stuart J., Diego M. Macall, and Peter W. B. Phillips. “Implications of Biological Information Digitization: Access and Benefit Sharing of Plant Genetic Resources.” *The Journal of World Intellectual Property* 23, no. 3–4 (2020): 267–287. Accessed June 25, 2025. <https://doi.org/10.1111/jwip.12151>

5 Pan-Canadian Genomics Strategy: What We Heard Report,” *Government of Canada*, 2023-03-13, <https://ised-isde.canada.ca/site/genomics/en/pan-canadian-genomics-strategy-what-we-heard-report#s5>

6 Smriti Mallapaty, “Scientists globally are racing to save vital health databases taken down amid Trump chaos”, *Nature*, 2025-02-07, <https://www.nature.com/articles/d41586-025-00374-y>



The \$175.1 million [Canadian Genomics Strategy](#), announced earlier this year, recognizes the central importance of data and makes improving data accessibility a core goal.⁷ Through the Strategy, Genome Canada and the Canadian Genomics Enterprise are investing in well-governed, interoperable and sovereign data assets.

Several other key projects are underway. These include Genome Canada's new landmark [Canadian Precision Health Initiative](#) (CPHI), which is building Canada's largest-ever collection of human genomic data—more than 100,000 genomes (over 20 petabytes of genomic data).

CPHI will provide the vast amounts of high-quality genomics data needed to develop AI-powered tools that enable more precise, preventative and cost-effective health care in Canada and give us a competitive edge in health innovation.

The [12 foundational CPHI research projects now underway](#) each have a specific goal for the data they will generate, from giving more Canadians with rare diseases a long-awaited diagnosis to improving treatment for cardiovascular disease. This work will begin advancing precision health care immediately, with data made accessible to researchers and industry via the [Pan-Canadian Genome Library](#) (PCGL).

The PCGL, supported by the Canadian Institutes of Health Research, is another essential, foundational project that will facilitate the intersection of AI and genomics. PCGL provides a unified, national approach for managing and sharing human genomic data in Canada while addressing jurisdictional and other fundamental considerations in the movement of human genetic data. These considerations include Indigenous data sovereignty and stewardship, and the ethical frameworks that uphold them, such as [The First Nations principles of OCAP®](#) (ownership, control, access, and possession).⁸

This work creates the platform for researchers and industry to use the vast trove of genomic data produced to facilitate breakthroughs and commercialized genomics solutions.

⁷ "Canadian Genomics Strategy," Government of Canada, 2025-02-27, <https://ised-isde.canada.ca/site/genomics/en/canadian-genomics-strategy>

⁸ The First Nations Information Governance Centre. "The First Nations Principles of OCAP®." *The Fundamentals of OCAP® Training*, accessed July 24, 2025. <https://fnigc.ca/ocap-training/>

What's next

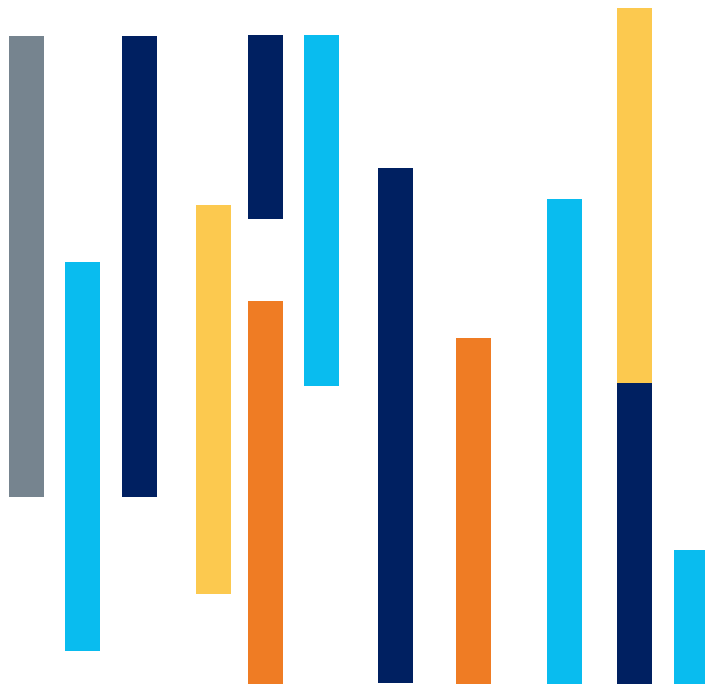
Understanding biological systems is central to addressing today's grand challenges, such as climate change, increasing food and energy demand, antimicrobial resistance and infectious disease. New growth opportunities across the bioeconomy—like synthetic biology, which builds and modifies biological systems so they can do new things or make specific products—will be transformative and deliver solutions previously thought impossible.

McKinsey Global Institute estimates that up to 60% of physical inputs to the global economy could soon be produced biologically, worth US\$2–4 trillion in direct economic impact by 2040.⁹

By leveraging the vast amounts of high-quality digitized biological data, AI is now helping genomics advance precision health, improve the productivity and resiliency of forests, crops and livestock, create more sustainable energy and biomaterials, bioremediate contaminated sites and protect the environment.

Canada's longstanding support for genomics and AI already leads to real-world innovation. However, as projects such as CPHI and PCGL reach their full fruition, Canada will have the sovereign data assets to accelerate this research to new levels.

In continuing to integrate AI and genomics and supporting the foundational work necessary to unleash new combinations of technologies and real-world innovations, Canada can stake out a stronger leadership role that will lead to economic and social benefits for years to come.



⁹ McKinsey & Company. "The Bio Revolution: Innovations Transforming Economies, Societies, and Our Lives." *McKinsey & Company*. Accessed July 24, 2025. <https://www.mckinsey.com/industries/life-sciences/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives>