La génomique, la numérisation de la biologie, et les possibilités que présente l'IA pour le Canada

Nous vivons actuellement une période de profonds bouleversements technologiques. Alors que les avancées dans les modèles de langage de grande taille de l'intelligence artificielle (IA) font les gros titres, une révolution tout aussi transformatrice — et étroitement liée — se déroule dans le domaine des sciences de la vie. La numérisation de la biologie grâce à la génomique modifie fondamentalement les fondements de secteurs entiers, des soins de santé à l'agriculture, en passant par les ressources naturelles et l'économie bleue. Pour exploiter pleinement le potentiel de la biorévolution alimentée par la génomique et l'essor de l'IA, il est nécessaire que ces technologies générales transformatrices se renforcent mutuellement.

La génomique est la science qui consiste à déchiffrer et à comprendre l'ensemble des informations génétiques d'un organisme codées dans l'ADN et les molécules apparentées. Cela inclut les êtres humains, les animaux, les plantes et les micro-organismes tels que les bactéries, les virus et les champignons.

Il a fallu environ 13 ans pour séquencer le premier génome humain, à un coût d'environ trois milliards de dollars américains. Aujourd'hui, le séquençage d'un génome humain peut être réalisé en une journée environ et ne coûte que quelques centaines de dollars, un chiffre qui ne cesse de baisser.[1] Ce rythme de progression à la baisse a largement dépassé la légendaire loi de Moore (voulant que la puissance de calcul des semi-conducteurs double tous les deux ans), soulignant *jusqu'à quel point* la recherche en génomique a été révolutionnaire. (Voir figure 1)

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

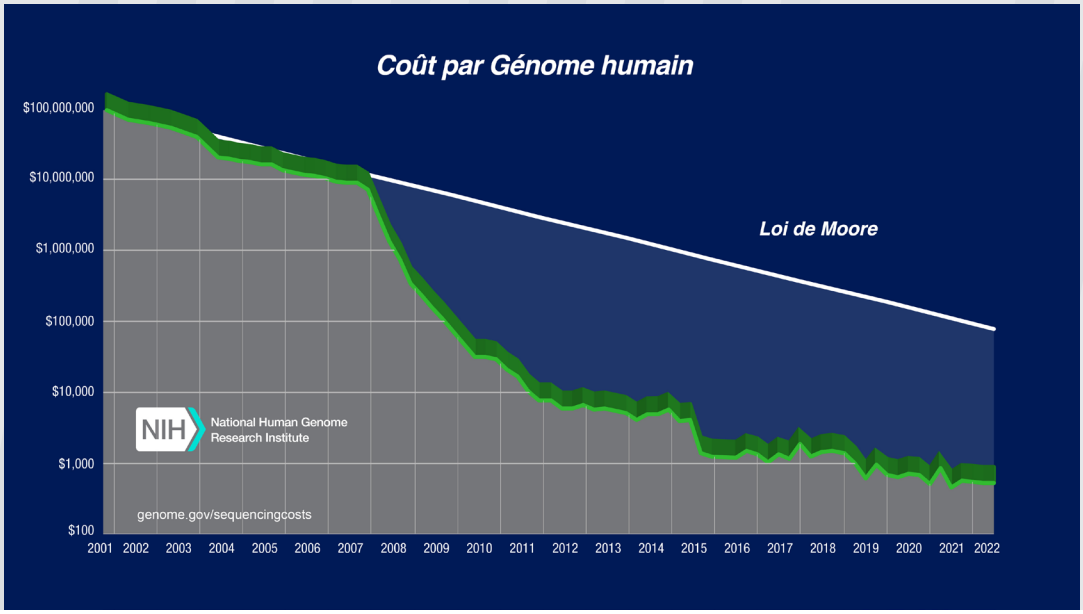


Figure 1 : Source : « DNA Sequencing Costs: Data », Kris A. Wetterstrand, « Coût du séquençage de l'ADN : données issues du programme de séquençage du génome du NHGRI, » National Human Genome Research Institute, 16-05-2023, https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data

Pour exploiter pleinement le potentiel de la génomique, il faut avant tout disposer de données. Il s'agit en effet d'un domaine qui nécessite un volume considérable de données. Un seul génome humain contient environ trois milliards de paires de bases d'ADN, et les données associées occupent entre 200 et 300 Go d'espace. La baisse rapide des coûts a entraîné une augmentation exponentielle du volume de données créées : des dizaines d'exaoctets (milliards de Go) par année rien que pour stocker les données génomiques *recueillies jusqu'à présent.*[2]

Cette numérisation de la biologie a alimenté l'utilisation de la génomique dans de nouveaux domaines de recherche et l'adoption rapide de la génomique clinique. Elle a mené au développement de technologies de pointe et d'innovations applicables dans tous les secteurs, notamment la santé publique, l'alimentation et les ressources naturelles. Les technologies issues de la génomique, telles que CRISPR et les vaccins à ARNm, sont tout simplement révolutionnaires.

Les investissements de longue date du Canada dans la recherche génomique dans divers secteurs — notamment la santé, l'agriculture, l'environnement, la foresterie, les pêches, l'énergie et l'exploitation minière — depuis la création de Génome Canada en 2000, lui confèrent un avantage stratégique sur bon nombre de ses homologues mondiaux qui ont principalement concentré leurs efforts en génomique dans le secteur de la santé.

Cependant, l'ampleur même de la production de données issues de la génomique pose des défis. Les chercheurs ont fait valoir que « l'accumulation rapide de données génomiques issues du séquençage de l'ADN et de données cliniques provenant des dossiers médicaux électroniques pose des défis et offre des possibilités considérables aux scientifiques des données pour extraire des informations biologiquement ou cliniquement pertinentes de la vaste quantité de données génotypiques et phénotypiques ».[3]

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

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

Le principal défi de la prochaine phase de la recherche génomique consiste à transformer ces données en informations exploitables qui permettront d'obtenir des résultats concrets dans le domaine des soins de santé et au-delà. C'est là qu'intervient la convergence de la génomique grâce à l'IA.

L'IA et la génomique

Compte tenu des énormes quantités de données produites par les chercheurs et les cliniciens, la génomique est de plus en plus étroitement liée à l'IA, chaque technologie contribuant à exploiter le potentiel de l'autre.

La plupart des recherches en génomique utilisent déjà l'IA. Cela inclut les projets dans lesquels :

- **l'IA est l'objectif qui définit le projet;**
- **l'IA est une activité qui contribue à l'objectif global du projet;**
- **l'IA est un outil qui est utilisé dans le cadre du projet.**

Génome Canada a financé son premier projet lié à l'IA en 2018 dans le cadre de son programme de bio-informatique et de génématique. Depuis, ce programme a financé plus de 75 projets, pour un investissement total de plus de 70 millions de dollars. Ces projets couvrent plusieurs secteurs et sont menés par des chercheurs de tout le pays représentant les six centres de génomique régionaux du Canada. Ils comprennent également des projets individuels menés par des chercheurs et des initiatives fondamentales majeures qui sont essentielles pour libérer le potentiel de la génomique grâce à l'utilisation de la génématique, de l'apprentissage automatique et de l'IA. Ensemble, ils ont positionné le Canada comme un acteur mondial de premier plan à la croisée de ces technologies.

Parmi les projets individuels tirant parti de la génématique, citons le travail de Fiona Brinkman, Gary Van Domselaar et William Hsiao, qui s'est appuyé sur l'apprentissage machine et la génomique pour améliorer les résultats sanitaires en conditions réelles. Ils ont développé une plateforme informatique, la « Integrated Rapid Infectious Disease Analysis (IRIDA) », afin de permettre une analyse plus génomique des éclosions de maladies infectieuses. Cet outil est désormais utilisé par le Laboratoire national de microbiologie du Canada et dans le cadre de collaborations internationales sur quatre continents.

Parallèlement, la recherche liée à l'IA soutient également d'autres domaines d'excellence industrielle au Canada, comme l'industrie laitière, qui représente 6,17 milliards de dollars. La génomique et l'IA sont mises à contribution pour produire des modèles prédictifs de la production laitière afin d'appuyer les pratiques de gestion et d'optimiser la rentabilité des fermes laitières dans le cadre d'un partenariat financé par Génome Canada entre Lactanet, My Intelligent Machines, et des équipes de recherche dirigées par Abdoulaye Baniré Diallo (Université du Québec à Montréal) et Marc-André Sirard (Université Laval).

Les projets de génomique assistée par l'IA débouchent également sur des technologies commercialisées et des entreprises dérivées. Creative Molecules Inc. [en anglais] est une entreprise dérivée du projet de Christoph Borcher et soutenu par Génome Canada, qui a permis de développer des algorithmes et des logiciels permettant d'interpréter les données de spectrométrie de masse afin de déterminer des modèles structurels en 3D. Ces travaux ont également mené à la création de trois serveurs Web qui mettent gratuitement à la disposition des chercheurs des outils de structure protéique, ce qui confère à cette recherche une plus grande valeur sociale.

À l'échelle mondiale, les bailleurs de fonds se sont concentrés sur l'activation du potentiel de l'IA et de la génomique. Sous l'administration précédente, l'agence National Institutes of Health des États-Unis a investi 130 millions de dollars américains dans le programme Bridge2AI, [en anglais] qui vise à accélérer l'utilisation de l'IA par les communautés de recherche biomédicale et comportementale. Genomics England a également lancé d'importants programmes de financement liés à l'IA, notamment un programme « Cancer 2.0 » de 26 millions de livres sterling [en anglais], qui vise à améliorer la précision et la rapidité du diagnostic des patients atteints de cancer grâce à l'évaluation de technologies de séquençage génomique de pointe utilisant l'IA pour analyser les données génomiques ainsi que des images numériques d'histopathologie et de radiologie.

Investir dans la souveraineté des données du Canada

Le Canada stimule également les progrès en génomique et en IA au-delà des projets de recherche individuels. De nouvelles initiatives importantes établissent les cadres de données et de gouvernance nécessaires pour rendre possible une révolution de la génomique alimentée par l'IA et pour fournir les données de haute qualité, diversifiées et fiables dont ont besoin les solutions d'IA porteuses d'avenir, alors que le Canada intensifie sa stratégie pancanadienne en matière d'intelligence artificielle.

Il est essentiel que les chercheurs et l'industrie aient accès à des données de qualité grâce à des modèles de gouvernance solides. La numérisation rapide des données biologiques remet déjà en question les modèles de gouvernance existants, tant au niveau national qu'international, étant donné que le rythme des progrès scientifiques dépasse de loin la capacité à réglementer.[4]

Il est de plus en plus reconnu qu'il est nécessaire de remédier à la fragmentation importante de l'écosystème de données du Canada et d'harmoniser les normes et les bases de données afin d'améliorer l'interopérabilité. Comme le gouvernement du Canada l'a appris lors de récentes consultations, les vastes quantités de données produites ont tendance à « résider dans des cloisonnements liés aux institutions et aux compétences, où les normes de données et les bases de données utilisées pour gérer les données ne sont pas harmonisées, et manquent d'interopérabilité ».[5] L'absence d'interopérabilité entre les données freine les applications possibles de l'IA pour soutenir la recherche en génomique.

Entre-temps, l'intensification des menaces géopolitiques et les réactions négatives à l'égard de la science à l'échelle mondiale ont également souligné l'importance de garantir l'accès aux données souveraines. Cette année, plusieurs collections de données fondamentales aux États-Unis, dont plusieurs bases de données essentielles sur la santé, ont été mises hors ligne.[6] Protéger la chaîne d'approvisionnement en données du Canada en fournissant un accès à des données organisées pertinentes pour les intérêts canadiens est une nécessité stratégique et essentielle pour réaliser le potentiel découlant de la numérisation de la biologie.

4 Smyth, Stuart J., Diego M. Macall, et 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. Consulté le 25 juin 2025. https://doi.org/10.1111/jwip.12151

5 "Stratégie pancanadienne en matière de génomique — Rapport « Ce que nous avons entendu » *Gouvernement du Canada*, 13-03-2023, https://ised-isde.canada.ca/site/genomique/fr/strategie-pancanadienne-matiere-genomique-rapport-nous-avons-entendu#s5

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

La Stratégie canadienne en matière de génomique de 175,1 millions de dollars, annoncée plus tôt cette année, reconnaît l'importance primordiale des données et fixe comme objectif fondamental d'en améliorer l'accessibilité.[7] Dans le cadre de cette stratégie, Génome Canada et l'Entreprise canadienne de génomique investissent dans des données bien gérées, interopérables et souveraines.

Plusieurs autres projets clés sont en cours. Parmi ceux-ci, mentionnons la nouvelle Initiative canadienne de santé de précision (ICSP) de Génome Canada, qui vise à constituer la plus grande collection de données génomiques humaines au Canada, soit plus de 100 000 génomes (plus de 20 pétaoctets de données génomiques).

L'ICSP fournira les vastes quantités de données génomiques de haute qualité nécessaires à la mise au point d'outils alimentés par l'IA qui permettront d'offrir des soins de santé plus précis, préventifs et rentables au Canada et nous donneront un avantage concurrentiel dans le domaine de l'innovation en santé.

Les 12 projets de recherche fondamentale menés actuellement dans le cadre de l'ICSP ont chacun un objectif précis en matière de données à produire, qu'il s'agisse de permettre à davantage de Canadiens atteints de maladies rares d'obtenir un diagnostic tant attendu ou d'améliorer le traitement des maladies cardiovasculaires. Ces travaux permettront de faire progresser immédiatement les soins de santé de précision, grâce au partage des données de la Bibliothèque génomique pancanadienne (BGP) avec les chercheurs et l'industrie.

La BGP, appuyée par les Instituts de recherche en santé du Canada, est un autre projet fondamental essentiel qui facilitera l'intersection entre l'IA et la génomique. La BGP offre une approche nationale unifiée pour la gestion et le partage des données génomiques humaines au Canada, tout en tenant compte des considérations juridictionnelles et autres considérations fondamentales relatives au transfert des données génétiques humaines. Ces considérations comprennent la souveraineté et la gestion des données autochtones, ainsi que les cadres éthiques qui les sous-tendent, tels que les principes de PCAP® des Premières Nations (propriété, contrôle, accès et possession).[8]

Ce travail crée une plateforme permettant aux chercheurs et à l'industrie d'utiliser la vaste mine de données génomiques produites afin de faciliter les percées et la commercialisation de solutions génomiques.

7 « Stratégie canadienne en matière de génomique », *Gouvernement du Canada*, 27-02-2025, https://ised-isde.canada.ca/site/genomique/fr/strategiecanadienne-matiere-genomique

8 Le Centre de gouvernance de l'information des Premières Nations. « Les principes de PCAP® des Premières Nations. » *Les cours Fondamentaux des principes de PCAP®*, consulté le 24 juillet 2025. https://fnigc.ca/ocap-training/.

Ce que l'avenir nous réserve

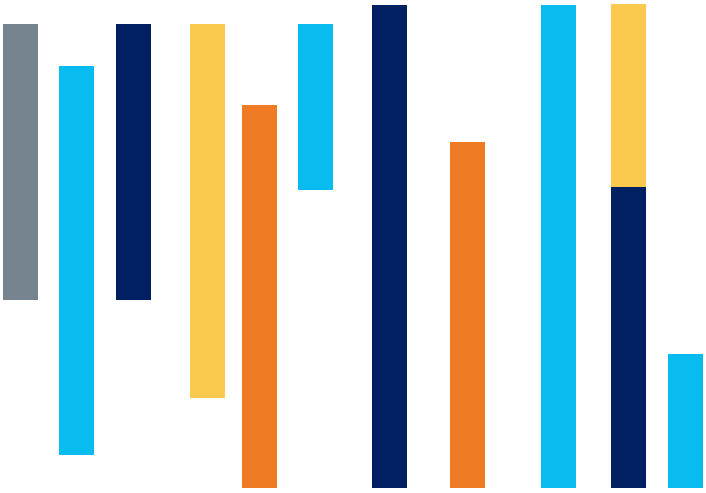
La compréhension des systèmes biologiques est essentielle pour relever les grands défis actuels, tels que le changement climatique, l'augmentation de la demande alimentaire et énergétique, la résistance aux antimicrobiens et les maladies infectieuses. De nouvelles possibilités de croissance dans le domaine de la bioéconomie — comme la biologie synthétique, qui construit et modifie des systèmes biologiques afin qu'ils puissent accomplir de nouvelles tâches ou fabriquer des produits spécifiques — seront transformatrices et apporteront des solutions jusque-là considérées comme impossibles.

La *McKinsey Global Institute* [en anglais] estime que jusqu'à 60 pour cent des intrants physiques de l'économie mondiale pourraient bientôt être produits biologiquement, ce qui représenterait un impact économique direct de 2 à 4 000 milliards de dollars américains d'ici à 2040.[9]

En exploitant les énormes quantités de données biologiques numérisées de haute qualité, l'IA aide désormais la génomique à faire progresser la santé de précision; à améliorer la productivité et la résilience des forêts, des cultures et du bétail; à créer des énergies et des biomatériaux plus durables; à assurer la bioremédiation des sites contaminés; et à protéger l'environnement.

Le soutien de longue date du Canada à la génomique et à l'IA a déjà donné lieu à des innovations concrètes. Cependant, à mesure que des projets tels que l'ICSP et la BGP atteindront leur pleine capacité, le Canada disposera des ressources souveraines en matière de données nécessaires pour accélérer cette recherche et la propulser vers de nouveaux sommets.

En poursuivant l'intégration de l'IA et de la génomique et en soutenant les travaux fondamentaux nécessaires pour exploiter de nouvelles combinaisons de technologies et d'innovations concrètes, le Canada peut se positionner comme un chef de file qui générera des avantages économiques et sociaux pour des années à venir.



9 McKinsey & Company. « The Bio Revolution: Innovations Transforming Economies, Societies, and Our Lives. » *McKinsey & Company.* Consulté le 24 juillet 2025. https://www.mckinsey.com/industries/life-sciences/our-insights/the-bio-revolution-innovations-transforming-economies-societies-and-our-lives.