

## AI in Focus: Genomics

September 28, 2021

By Tony Orsi and Iris Cheung

Genomics is the study of our genetic material. In the clinic, genomic data can be used to guide diagnostic and therapeutic decision-making. The concept that the effectiveness of a treatment may be determined by a patient's genotype has a long history. However, the adoption of genome-level information in medicine has only become possible with the development of next-generation sequencing (NGS) technologies and the expansion of computational infrastructure, making the "\$1,000 genome" a reality. Further, the latest advance in sequencers has also driven down the average time to sequence one human genome to as little as an hour. While the cost of generating genomic data is no longer prohibitive, the analysis and interpretation of the vast amount of data generated by these sequencing technologies and the use of this data for clinical applications remains a challenge.

Some challenges are inherent in the technologies themselves. For example, nanopore-based sequencing technologies rely on the changes in an electrical current as a piece of nucleic acid passes through a membrane via a nanopore. While nanopore sequencing technology provides advantages such as long sequencing reads and direct detection of nucleic acid modifications, its error rate is relatively high. However, with the advent of AI, deep learning algorithms have been developed to improve raw read accuracy and to correct errors post-basecalling.

Other challenges are inherent in the nature of the samples. Tumor samples, for example, can be highly heterogeneous, harboring many somatic mutations that occur at low frequencies. This complicates the process of variant calling and requires refinement to remove false positives. However, it has been demonstrated that machine learning can be applied to automate the refinement step of variant calling for cancer sequencing data, which otherwise would require manual review of the aligned reads.

Interpretation of a variant can take on different directions, depending on the type of the variant (such as a single nucleotide variant (SNV), an insertion/deletion, etc.) and the location of the variant. For instance, if an SNV is located within a coding region leading to an amino acid change, understanding how the protein structure may be impacted could shed some insight into a disease mechanism. However, deep neural networks, which are a powerful form of AI, have recently been used to develop a 3D protein modeling method that has demonstrated the ability to predict the structure of a protein from its genetic sequence at a much higher accuracy compared to other methods. When a variant is located in a non-coding region, the variant may still exert biological consequences such as, for example, by affecting DNA modification or binding of transcription factors. De novo sequence-based prediction of non-coding variant effects based on deep learning has been reported to predict certain chromatin features with high accuracies. The framework has been further developed to predict ab initio the effects of variants on gene expression levels.

As mentioned above, genomic medicine involves the use of genomic data in diagnostic and therapeutic decision-making. There are now many examples of rare and undiagnosed diseases being diagnosed with exome or genome sequencing. For example, in an earlier study, exome sequencing was used in a patient with a Crohn disease-like condition, but no definitive diagnosis could be arrived at based on conventional clinical evaluation. However, exome sequencing revealed a novel mutation in a gene involved in the inflammatory response and programmed cell death but was not previously associated with Crohn disease. The diagnosis led to an effective treatment.

More recently, an active area of research is to integrate data from omics with clinical and environmental data. Due to the complexity and the amount of data involved, the use of AI tools has enabled determining patterns from these disparate types of data and making predictions therefrom. For example, checkpoint inhibitors are very effective for late-stage cancers



for certain patients, but response rates vary across patients. However, machine learning algorithms trained on data derived from whole exome sequencing, RNA-Seq, and clinical features have been used successfully to predict patient response to a checkpoint inhibitor immunotherapy. In addition, integrating genomic and environmental exposure data collected through wearable biosensors using machine learning methods can improve our understanding of the complexities in gene-environment interactions [3] and has potential applications in health management.

Patent trends can be used to provide insights into commercial activities in specific fields or sectors. An analysis of patent filings relating to AI and genomics shows that in the US, the number of patent applications yearly has more than doubled since 2015 (Figure 1) [4] with a similar trend seen in Canada (Figure 2). The overall trend points to an accelerating adoption of AI technologies in commercial products and services relating to health.

Figure 1. Application trend in the US

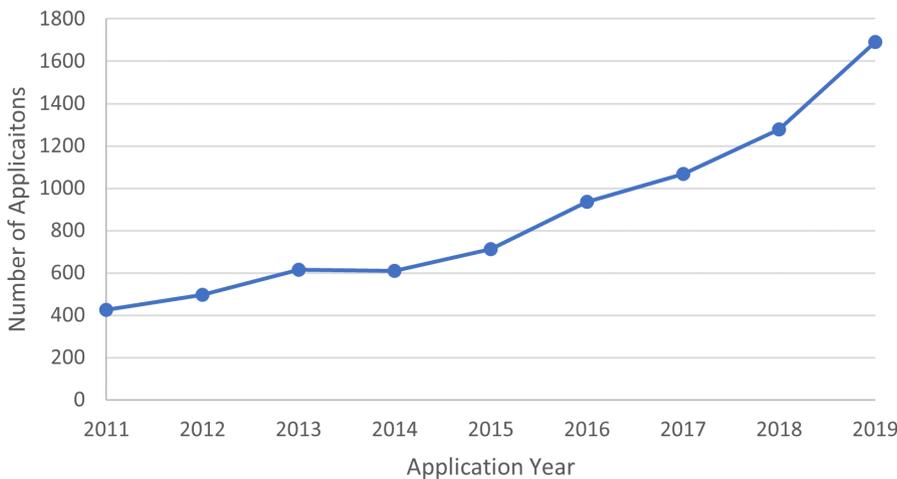
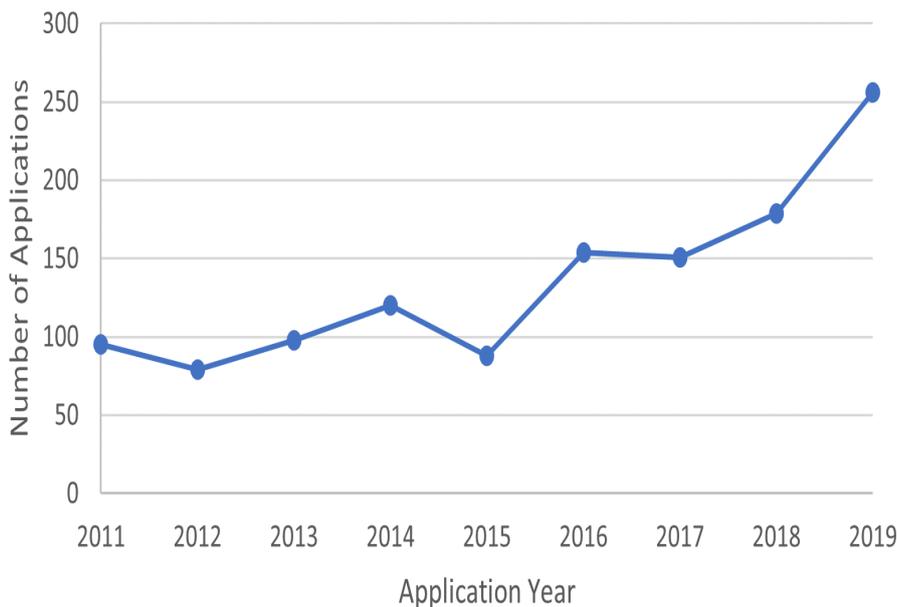


Figure 2. Application trend in Canada



In Canada, recent developments in patent law have the potential to provide a more favorable environment for patenting of these



technologies. Specifically, *Yves Choueifaty v Attorney General of Canada*<sup>[15]</sup> and the resulting practice notice on patentable subject matter issued by CIPO are improving the chances that medical diagnostic methods are patentable subject matter in Canada<sup>[6]</sup>. These developments can create opportunities for companies and other organizations to capture value through patent protection thereby creating a virtuous cycle for developing further diagnostic tools that may increase the adoption of omics and other types of big data platforms in medicine and the consumer wellness space.

Accordingly, looking forward, we expect to see a continued increase in innovation for AI tools that may be developed to understand disease mechanisms, discover therapeutic targets, or evaluate treatment outcome for the benefit of individualized patient care. For those interested in protecting such innovations to improve their ROI and build value, please feel free to contact one of our AI practice group members.

This has been the seventh article in our *AI in Focus* series. You can read the first six articles here:

- [AI in Focus – Autonomous Vehicles](#)
- [AI in Focus – Fundamental Artificial Intelligence and Video Games](#)
- [AI in Focus – Robotics](#)
- [AI in Focus – Natural Language Processing](#)
- [AI in Focus – BlueDot and the Response to COVID-19](#)
- [AI in Focus – Image Recognition](#)

If you have any ideas for other topics that you would like us to cover in our next article in this series, please email [Isi Caulder](mailto:lsi@bereskinparr.com), Co-Leader of the Artificial Intelligence (AI) practice group at Bereskin & Parr LLP.

---

[1] <https://hdr.mitpress.mit.edu/pub/y7r65r4k/release/2>

[2] <https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost>

[3] <https://www.sandiegouniontribune.com/business/biotech/sd-me-illumina-novaseq-20170109-story.html>

[4] <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-019-1727-y>

[5] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7434944/>

[6] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6428590/>

[7] <https://www.nature.com/articles/d41586-020-03348-4>

[8] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4768299/>

[9] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6094955/>

[10] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6929710/>

[11] <https://www.nature.com/articles/gim9201146>

[12] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6312275/>

[13] <https://doi.org/10.1016/j.cell.2019.03.004>

[14] We analyzed patent application trends using the PatSnap® database and a search strategy that was adapted from



WIPO Technology Trends 2019 Artificial Intelligence

([https://www.wipo.int/export/sites/www/tech\\_trends/en/artificial\\_intelligence/docs/techtrends\\_ai\\_methodology.pdf](https://www.wipo.int/export/sites/www/tech_trends/en/artificial_intelligence/docs/techtrends_ai_methodology.pdf))

[15] <https://www.canlii.org/en/ca/ct/doc/2020/2020fc837/2020fc837.html?resultIndex=1>

[16]

<https://www.bereskinparr.com/doc/the-canadian-patent-office-releases-new-guidelines-the-diagnostics-industry-should-be-cautic>

[17] <https://www.bereskinparr.com/practicearea/artificial-intelligence-ai>

Content shared on Bereskin & Parr's website is for information purposes only. It should not be taken as legal or professional advice. To obtain such advice, please contact a Bereskin & Parr LLP professional. We will be pleased to help you.