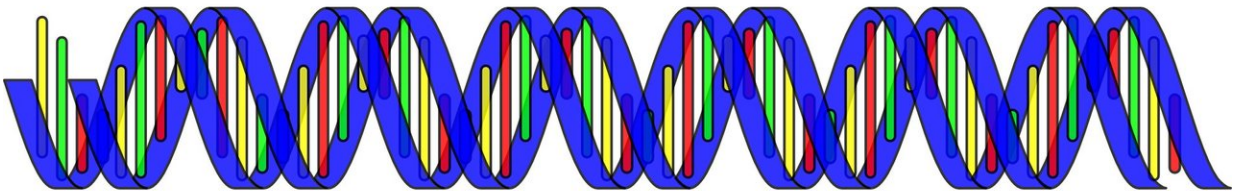# Connecting the dots for health data

November 18 2021

Credit: CC0 Public Domain

Precision medicine requires big data. In order to improve the treatment of individuals with cancer, or to understand rare diseases, scientists and clinicians, as well as AI technologies require access to larger sets of health research data that covers diverse populations and wide ranges of conditions. For AI, more data means a better understanding of diseases, which will lead to more accurate diagnosis and treatment. At the same time, each hospital will only see a relatively small number of individuals with a disease, and even across the province, we have access to only a small portion of the total data available worldwide. To build the large-

scale datasets needed to drive forward precision medicine, sharing of data across the country and around the world is critical.

The Canadian Distributed Infrastructure for Genomics (CanDIG), featured recently in a special issue of *Cell Genomics* dedicated to data sharing, is Canada's solution to enabling data sharing across the country (and connecting our data to datasets around the world). Led out of Toronto's University Health Network with sites at Montreal's McGill University and BC Genome Science Center, CanDIG is a collaboration of computer scientists, AI specialists, clinicians, and geneticists working together to enable studies needed to address the health challenges faced by Canadians.

CanDIG is a driver project of the Global Alliance for Genomics and Health (GA4GH), an [international effort](link) setting standards for genomics and health data looking to improve interoperability across the genomics landscape worldwide. The organization served as the focus for this month's special issue of *Cell Genomics* for their work on global genomics and health data sharing efforts. Canada has been a leader in the GA4GH, hosting its headquarters, leading several project work-streams, and implementing many GA4GH standards. CanDIG, as one of the driver projects, has not only implemented GA4GH standards but helped inform and build many of them. CanDIG is already helping scientists nationally access large-scale genomics data that was previously siloed in individual provinces or hospitals and is starting to connect Canada's genomic datasets to those from around the world through collaborations such as the EU/Africa/Canada CINECA project.

The CanDIG platform was developed to address Canada's province-based healthcare and privacy legislation, building a federation of datasets, simplifying the challenges of sharing across provincial borders. CanDIG is also a key component of the upcoming Digital Health and Discovery Platform (DHDP), a $200 million dollar effort funded in part

by the Canadian government, which will support sharing of genomic data from the Terry Fox Marathon of Hope Cancer Centres Network. Making this data available to researchers is key to unlocking its potential for discovery, and enabling better cancer treatments, because the smartest researchers and the most powerful machine learning techniques can't do anything with data they can't find, access or use.

"At institutions like UHN, we're building increasingly sophisticated data resources containing health data from many different sources. The next step is to help researchers turn that data into new knowledge by making it findable, available and usable in a uniform, curated, secure way, and allowing it to be combined with similar datasets across other hospitals. CanDIG is a significant step towards enabling researchers from across Canada to access the wealth of data being collected and generated anywhere in the country," says Dr. Michael Brudno, CanDIG Principal Investigator, UHN's Chief Data Scientist, and Professor of Computer Science at the University of Toronto.

"Participating within the GA4GH community and international projects like the EU/Africa/Canada CINECA project, CanDIG is starting to connect Canadian genomics efforts to those around the world. As health data types grow richer and volumes increase we need to make sure our datasets are findable and useful; Canada is a world leader in this," explains Dr. Guillaume Bourque, CanDIG Co-lead, Professor of Molecular Genetics at McGill University & Director of the Canadian Center for Computational Genomics (C3G).

"Access to whole-genome data has been vital to understanding the spectrum of mutations that accrue in cancer. CanDIG and the Terry Fox Digital Health and Discovery Platforms will help the data collected by the Marathon of Hope Cancer Centres Network be studied by as many approved researchers as possible," concludes Dr. Steve Jones, CanDIG Co-lead, Head of Bioinformatics and Co-Director at the BC Michael

Smith Genome Sciences Centre.

Provided by University Health Network

Citation: Connecting the dots for health data (2021, November 18) retrieved 4 May 2024 from https://medicalxpress.com/news/2021-11-dots-health.html