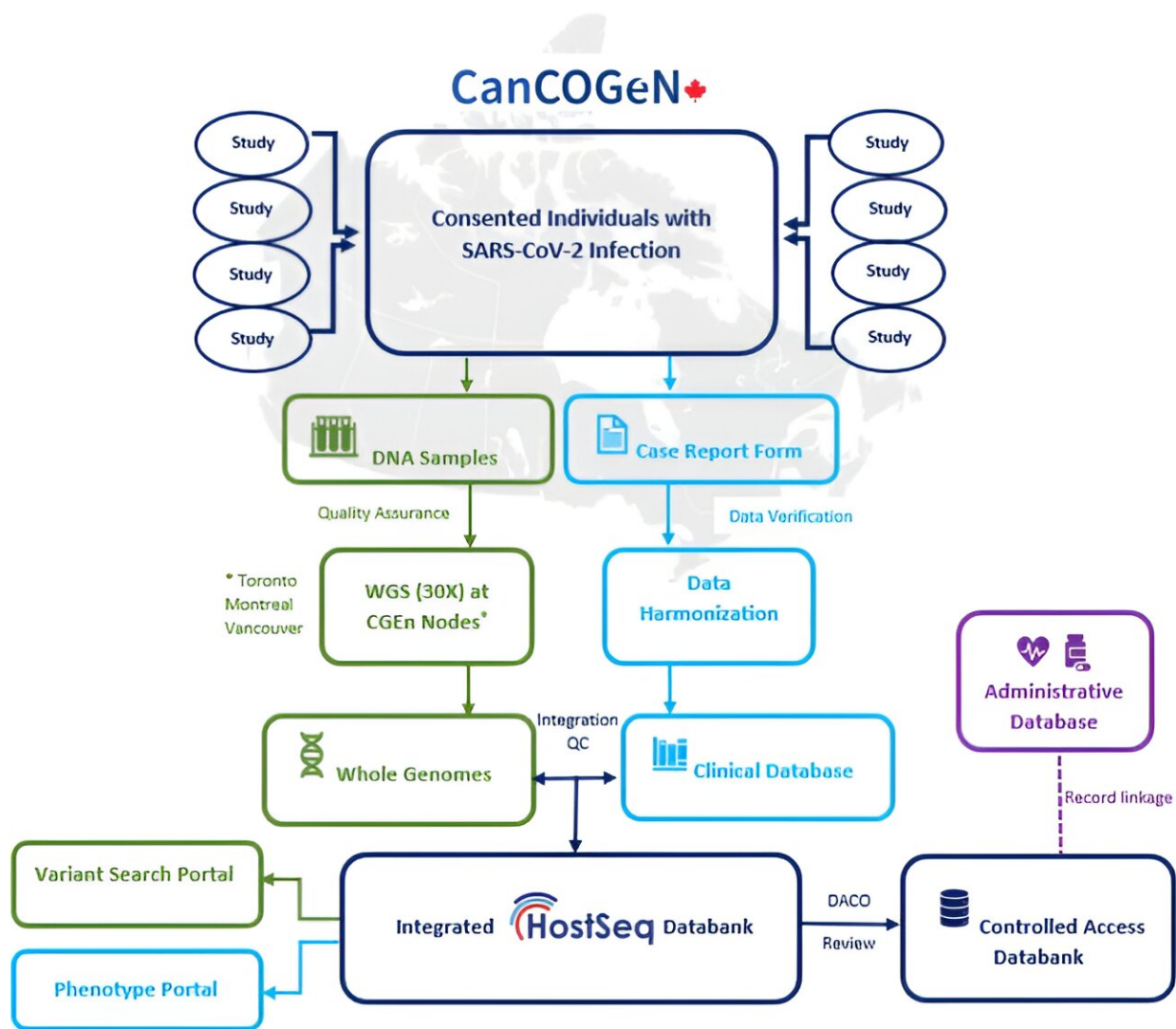


Why is COVID-19 more severe in some people? Researchers use genetics, data science to find out

October 25 2023, by Tyler Irving



Sample and data flow in HostSeq (Aspects of graphics acquired from Wikimedia)

Commons). Credit: *BMC Genomic Data* (2023). DOI: 10.1186/s12863-023-01128-3

Why do some people have a more severe course of COVID-19 disease than others? A genome sequence database created by an international collaboration of researchers, including many from the University of Toronto and partner hospitals, may hold the answers to this question—and many more.

The origins of the Canadian COVID-19 Human Host Genome Sequencing Databank, known as CGEn HostSeq, can be traced to the earliest days of the pandemic.

Lisa Strug, senior scientist at The Hospital for Sick Children (SickKids) and academic director of U of T's Data Sciences Institute, one of several U of T institutional strategic initiatives, says [genetic data](#) was top of mind for her and other researchers in late 2019 and early 2020 as reports of a novel form of coronavirus emerged from China and then other locations across the globe.

"In my research, I use data science techniques to map the genes responsible for complex traits," says Strug, who is a professor in U of T's departments of statistical sciences and computer science in the Faculty of Arts & Science and in the biostatistics division of the Dalla Lana School of Public Health.

"We knew that genes were a factor in the severity of previous SARS infections, so it made sense that COVID-19, which is caused by a closely related virus, would have a genetic component, too.

"Very early on, I started getting messages from several scientists who

wanted to set up different studies that would help us find those genes."

Over the next few months, Strug—who is also the associate director of SickKids' Centre for Applied Genomics, one of three sites across Canada that form CGEn, Canada's national platform for genome sequencing infrastructure for research—collaborated with nearly 100 researchers from across U of T and partner hospitals and institutions, as well as other researchers from across Canada to enroll individuals with COVID-19 and sequence their genomes.

Some of the key team members from the Toronto community included: Stephen Scherer, chief of research at SickKids Research Institute and a University Professor in U of T's Temerty Faculty of Medicine, as well as director of the U of T McLaughlin Centre; Rayjean Hung, associate director of population health at the Lunenfeld-Tanenbaum Research Institute, Sinai Health, and a professor in U of T's Dalla Lana School of Public Health; Angela Cheung, clinician-scientist at University Health Network, senior scientist at Toronto General Hospital Research Institute and a professor in U of T's Temerty Faculty of Medicine; and Upton Allen, head of the division of infectious diseases at SickKids and a professor in U of T's Temerty Faculty of Medicine.

Identifying associations between individual genes and complex traits typically requires thousands of genomes—both from those with the trait and those without. Though there was no shortage of cases to choose from, it was critical to gather and sequence DNA—and then organize the data in a way that would be ethical, efficient and useful to researchers now and in the future.

"One of our key mandates at the Data Sciences Institute is developing techniques and programs that ensure that data remains as open, accessible and as re-producible as it can be," Strug says.

"That vision was brought to bear as we assembled the data infrastructure for this project—for example, ensuring that consent forms were as broad as possible so that this data could be linked with other sources, from electronic medical records to other health databases.

"We wanted to be sure that even after the COVID-19 pandemic was over this could be a national whole genome sequencing resource to ask all kinds of questions about health and our genes. The development of the database and its open nature also enabled Canada to collaborate effectively with similar projects in other countries."

In the end, [the project gathered more than 11,000 full genome sequences from across Canada](#), representing patients with a wide range of [health outcomes](#). Results are published in *BMC Genomic Data*. Those data were then combined with even more sequences from patients in other countries under what came to be called the COVID-19 Host Genetics Initiative.

It didn't take long for patterns to start to emerge. [A paper](#) published in *Nature* in 2021 identified 13 genome-wide significant loci that are associated with SARS-CoV-2 infection or severe manifestations of COVID-19.

Since then, even more data have been added, and subsequent analysis has confirmed the significance of existing loci while also identifying new ones. The [most recent update](#) to the project, published in *Nature* earlier this year, brings the total number of distinct, genome-wide significant loci to 51.

"Identification of these loci can help one predict who might be more prone to a severe course of COVID-19 disease," says Strug.

"When you identify a trait-associated locus, you can also unravel the

mechanism by which this genetic region contributes to COVID-19 disease. This potentially identifies therapeutic targets and approaches that a future drug could be designed around."

While it will take many more years to fully untangle the effects of the different loci that have been identified, Strug says that the database is already showing its worth in other ways.

"It can be difficult to find datasets with whole genome sequence and approved for linkage with other [health information](#) that are this large, and we want people to know that it is open and available for all kinds of research well beyond COVID through a completely independent data access committee," she says.

"For example, several investigators from across Canada have been approved to use these data and we've even provided funding to trainees to encourage them to develop new data science methodologies or ask novel health questions using the CGen HostSeq data.

"This was a humongous effort, where researchers from across Canada came together during the COVID-19 pandemic to recruit, obtain and sequence DNA from more than 11,000 Canadians in a systematic, co-operative, aligned way to create a made-in-Canada data resource that will hopefully be useful for years to come. I think that was really miraculous."

More information: S Yoo et al, HostSeq: a Canadian whole genome sequencing and clinical data resource, *BMC Genomic Data* (2023). [DOI: 10.1186/s12863-023-01128-3](https://doi.org/10.1186/s12863-023-01128-3)

Provided by University of Toronto

Citation: Why is COVID-19 more severe in some people? Researchers use genetics, data science to find out (2023, October 25) retrieved 27 April 2024 from <https://medicalxpress.com/news/2023-10-covid-severe-people-genetics-science.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.