

ICGC brings more genomic health data to researchers on the Amazon Web Services Cloud

November 18 2015

The International Cancer Genome Consortium (ICGC) announced today that 1,200 encrypted cancer whole genome sequences are now securely available on the Amazon Web Services (AWS) Cloud for access by cancer researchers worldwide.

The Ontario Institute for Cancer Research (OICR), which houses the ICGC's Data Coordination Center (DCC), copied ICGC genome data onto the AWS Cloud and is providing authorized researchers with credentials to access and analyze the data using secure mechanisms. The ICGC Data Access Compliance Office has established a framework that protects the confidentiality of research participants while working to ensure that the research will benefit future [cancer patients](#).

The newly launched initiative means one of the world's largest collections of cancer genome data is now more easily accessible to qualified researchers, which will enhance collaboration and potentially accelerate the development of new treatments for cancer patients.

Cloud solutions have become essential to genomics research because of the vast amount of data produced by researchers and the difficulties inherent in transferring such large datasets between sites. Projects can quickly grow to several petabytes in size, with each petabyte being the equivalent of data on 223,000 DVDs. Very few institutions around the world have the capacity to download such immense datasets for analysis,

and this has limited the number of researchers who can access genome projects and the scope of what can be done with the data.

With cloud computing, researchers don't need to download data. They can work with data and run experiments in the cloud, a flexible network of servers on the Internet, and access data in minutes rather than months. Data stored in the cloud has been shown to be as secure, if not more so, than data downloaded to local servers and hard drives. The set of 1,200 genomes now available on AWS is the first installment of ICGC data to be posted and is expected to grow several fold over the next 12 months with the addition of data from more cancer patients.

"This initiative brings together one of the world's largest cancer genome datasets and one of the world's leading cloud computing providers to create a powerful new resource for cancer researchers," said Dr. Lincoln Stein, Director of the Informatics and Biocomputing Program at the Ontario Institute for Cancer Research and Director of the ICGC's Data Coordination Center. "Now, far more researchers will have access to ICGC data, opening up the possibility of new discoveries and new breakthroughs in [cancer research](#)."

The Pan-Cancer Analysis of Whole Genomes (PCAWG) project of the ICGC and The Cancer Genome Atlas (TCGA) is coordinating analysis of more than 2,800 cancer genomes, and is making extensive use of AWS and the genomes stored on Amazon Simple Storage Service (Amazon S3). Each genome is being characterized through a suite of standardized algorithms, including alignment to the reference genome, uniform quality assessment, and the calling of multiple classes of somatic mutations. Scientists participating in the research projects of PCAWG are addressing a series of fundamental questions about cancer biology and evolution based on these data.

"Making this data available and usable will enable more researchers

across the world to ask questions and get answers that were previously out of reach," said Matt Wood, General Manager of Product Strategy at Amazon Web Services, Inc. "Researchers can now explore these large and diverse datasets in unconstrained ways, without having to manage large amounts of physical infrastructure. Instead, they can focus on driving their state-of-the-art research forward."

"Cancer research is becoming increasingly data-heavy. Compiling the data, organizing the data, analyzing the data, making the data available to all researchers—these are fundamental to making further progress in cancer genome research, and we are excited at the possibilities of working with innovative cloud-based computing systems to achieve these advances," said Peter Campbell, Head of Cancer Genetics and Genomics at the Wellcome Trust Sanger Institute, who is helping to lead the PCAWG project.

"In the next year, it is estimated that 14 million people worldwide will learn that they have cancer. In order to accelerate our understanding of this disease and ultimately provide better treatment, it is critical that we develop solutions able to meet the scale of this challenge. Co-localizing ICGC data as well as other cancer genomics data sets like The Cancer Genome Atlas with secure and scalable computation resources represents a major step forward for both researchers and patients. With ICGC data available on AWS, we utilized the Seven Bridges platform to perform variant calling on hundreds of genomes weeks faster than would have been possible using local infrastructure," said Deniz Kural, CEO of Seven Bridges Genomics and Principal Investigator of one of three NCI-funded Cancer Genomics Cloud pilot projects.

"This effort to provide the ICGC datasets on AWS will lower the barriers currently associated with computing on thousands of genomes. Users will have the ability to quickly analyze datasets within the cloud on highly scalable infrastructure. This is a paradigm shift from the old

model of slowly downloading data to a user's local infrastructure before any meaningful work can commence," said Brian O'Connor, Managing Director of Cloud Computing at the Ontario Institute for Cancer Research.

"The ICGC Data Access Compliance Office (DACO) has been a forerunner in providing controlled, secure, and efficient access to cancer genomic data to members of the research community. It now welcomes the opportunity to further advance research for the benefit of all cancer patients by enabling controlled cloud access to ICGC genomic data stored on AWS. Throughout the process, DACO will implement a robust governance framework to ensure a high degree of privacy protection to patients' genetic and health [data](#)," said Yann Joly, Data Access Officer, ICGC DACO, McGill University.

"This exciting collaboration and new use for cloud technology is the future of cancer research. Ontario is proud to be part of this initiative through the Ontario Institute for Cancer Research and we look forward to seeing this relationship help [cancer](#) patients around the world," said Reza Moridi, Ontario's Minister of Research and Innovation.

Provided by Ontario Institute for Cancer Research

Citation: ICGC brings more genomic health data to researchers on the Amazon Web Services Cloud (2015, November 18) retrieved 24 April 2024 from <https://medicalxpress.com/news/2015-11-icgc-genomic-health-amazon-web.html>

<p>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.</p>
--