

Genomic Data Commons at University of Chicago launches new era of cancer data sharing

6 June 2016

The Genomic Data Commons (GDC), a next-generation platform that enables unprecedented data access, analysis and sharing for cancer research, publicly launched at the University of Chicago on June 6, opening the door to discoveries for this complex set of diseases.

The GDC went live with approximately 4.1 petabytes of data from National Cancer Institute-supported research programs, including some of the largest and most comprehensive cancer genomics datasets in the world—such as The Cancer Genome Atlas and Therapeutically Applicable Research to Generate Effective Treatments—and more than 14,000 anonymized patient cases. One petabyte equals 1 million gigabytes.

Vice President Joe Biden toured the GDC operations center at the University of Chicago in advance of his appearance to announce the project at the annual meeting of the American Society for Clinical Oncology on June 6.

The Data Commons centralizes, standardizes and harmonizes genomic and [clinical data](#) on a unified and interoperable platform. Cancer researchers can access these data for analyses and submit their own datasets to share with the research community. By making high-quality data broadly accessible, the GDC provides much-needed tools to accelerate studies of the biological mechanisms of cancer and the development of personalized treatments for individual patients.

UChicago developed and operates the Data Commons with NCI funding under a subcontract from Leidos Biomedical Research at the Frederick National Laboratory for Cancer Research, in collaboration with the Ontario Institute for Cancer Research.

Development of the GDC began in 2014 at UChicago's Center for Data Intensive Science (CDIS). Over the past two years, the team has created an innovative suite of tools, software and infrastructure—based on CDIS open-source projects such as the Bionimbus Protected Data Cloud—to curate the massive amounts of data held by the GDC.

"Today, making discoveries from cancer [genomic data](#) is challenging because diverse research groups analyze different cancer datasets using various methods that are not easily comparable," said GDC principal investigator Robert Grossman, professor of medicine and director of CDIS at UChicago. "The Genomic Data Commons brings together genomic datasets and analyzes the data using a common set of methods so that researchers may more easily make discoveries, and, in this sense, democratizes the analysis of large cancer genomic datasets."

"Big data" is recognized as essential to efforts in understanding and treating cancer. Cancer is as complex as it is devastating. It involves a host of genetic, lifestyle and environmental factors, and is now known to comprise hundreds of diseases—each with unique features, driving forces and vulnerabilities to treatments. Large sample sizes are required to provide the statistical power to understand which combinations of drugs are effective against which combinations of mutations that drive cancer.

Breaking Barriers

While enormous amounts of genomic and clinical data have been gathered by NCI-funded research, several barriers have prevented researchers from making full use of them. Genomic data from different projects, clinical trials and cancer types

are siloed in different locations with local management systems, making data sharing difficult. These large datasets can take months to download, and not all researchers have access to the sophisticated tools needed to study them. In addition, disparate collection and analysis approaches by separate research groups inhibit collaborative work.

The GDC breaks down these barriers by bringing cancer genomics datasets and associated clinical data into one location that any researcher may access. It harmonizes the data with a common set of analytic pipelines to make it easier to study the information, which in the past has typically been available as separate datasets analyzed with separate pipelines. By making these data available using modern computing and network technology, the GDC makes it possible for any researcher to ask new and fundamental questions about cancer.

Built and managed by Grossman's team at the University of Chicago, the GDC will:

- Serve as a central unified repository for cancer genomic data and associated clinical data.
- Clean, standardize and harmonize data, as well as provide quality control, so that analyses can be conducted using common algorithms and pipelines.
- Support basic research and clinical trials by making data easily accessible, findable, interoperable and reusable.
- Provide powerful data transfer, search, Application Programming Interface (API) and analysis tools to researchers at no cost.

tools needed to share and study them, and support to submit their own data. It will house data from a new era of programs that will sequence the DNA of patients enrolled in NCI clinical trials. These datasets will lead to a much deeper understanding of which therapies are most effective for different cancers. The GDC will support clinical trials that focus on single patients, known as "n of 1" [clinical trials](#), and will become an important component in how precision medicine is used to treat individual patients.

The GDC also creates a foundation for future cloud-based technologies that could allow researchers to analyze large-scale datasets and perform experiments remotely, such as through the NCI's Cancer Cloud Pilots Program. In addition, the open-source software being developed by the CDIS has the potential to become a model for data-intensive research efforts for other diseases, such as Alzheimer's and diabetes, which would greatly benefit from similar large-scale, data-driven approaches to develop cures.

"We are at a crossroads today in whether we will have the critical mass of cancer-related data needed to power new discoveries and improve [cancer](#) care," Grossman said. "Over time, I expect the GDC will play a more and more important role in providing the data required at the scale required so that precision medicine fulfills its promise."

Provided by University of Chicago Medical Center

A Foundation for the Future

As the first step in a next generation knowledge system for cancer, the GDC enables and accelerates efforts to identify both high- and low-frequency cancer driver mutations, assists in revealing the genetic determinants of response to therapy, and informs the composition of clinical trial cohorts.

The GDC will help bridge siloes by providing researchers with access to high-quality data, the

APA citation: Genomic Data Commons at University of Chicago launches new era of cancer data sharing (2016, June 6) retrieved 19 August 2022 from <https://medicalxpress.com/news/2016-06-genomic-commons-university-chicago-era.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.