

# University of Chicago to establish Genomic Data Commons

2 December 2014

The University of Chicago is collaborating with the National Cancer Institute to establish the nation's most comprehensive computational facility that stores and harmonizes cancer genomic data generated through NCI-funded research programs.

The establishment of the NCI Genomic Data Commons (GDC) will expand access for scientists around the country, speeding up research and, in turn, leading to faster discoveries for patients. The GDC will provide an interactive system for researchers, making the data easier to use; it also will provide resources to facilitate the identification of subtypes of [cancer](#) as well as [potential therapeutic targets](#).

"The Genomic Data Commons has the potential to transform the study of cancer at all scales," said Robert Grossman, Ph.D., director of the GDC project and professor in the Department of Medicine at the University of Chicago. "It supplies the data so that any researcher can test their ideas, from comprehensive 'big-data' studies to genetic comparisons of individual tumors to identify the best potential therapies for a single patient."

NCI has funded a number of large research projects that have collected [genomic data](#) on tumor types from more than 10,000 patients. However, the data for these studies are scattered across different locations and are in different formats, making it challenging for researchers to perform analyses. As genome sequencing technology continues to evolve and datasets become increasingly larger and more complex, this situation will get more problematic. According to an Institute of Medicine report, there is an urgent need for a system to store, harmonize and analyze existing cancer genomics data, which currently amounts to roughly 20 petabytes of information - 10 times as much as all of the publications currently housed in U.S. academic research libraries.

## Data Democracy

To address these challenges, the GDC will provide an expandable, modern informatics framework that uses standards to make raw and processed genomic data broadly accessible. The GDC will harmonize and centralize existing NCI datasets through an approach to data storage and analysis similar to what is used by companies such as Google and Facebook. The GDC will eliminate a major chokepoint, streamlining access to data for researchers regardless of their institution's size or budget—effectively democratizing access to the material. It will also enable previously unfeasible collaborative efforts between scientists.

"With the GDC, the pace of discovery shifts from slow and sequential to fast and parallel," said Conrad Gilliam, Ph.D., dean for basic science at the University of Chicago Biological Sciences Division. "Discovery processes that today would require many years, millions of dollars, and the coordination of multiple research teams could literally be performed in days, or even hours."

The GDC serves as a key step toward the development of precision medicine—targeted treatments that are tailored to individual patients. Once fully developed, it will provide an interactive system for researchers and clinicians to upload their cancer genomics data and use it to identify the molecular subtype of cancer and potential therapeutic targets. Genetic data will be linked to extensive clinical information from patients and their response to treatment.

"The availability of high-quality genomic data and associated clinical annotations is extremely important because this information can be combined and mined repeatedly to make new discoveries," said Louis Staudt, Ph.D., M.D., director of NCI's Center for Cancer Genomics.

## Foundation for the Cloud

The GDC also creates a foundation for future cloud-based technologies that one day will allow researchers to analyze large-scale datasets and perform experiments remotely. The open-source software being developed by the GDC has the potential to become a model for data-intensive research efforts for other diseases such as Alzheimer's and diabetes, which desperately need similar large-scale, data-driven approaches to develop cures.

"The GDC is absolutely needed," said Jean Zenklusen, M.S., Ph.D., director of The Cancer Genome Atlas (TCGA) program office at NCI. "The current scale of the data is such that mostly big institutes with large bioinformatics cores are the only ones who have been able to take advantage of the huge amount of [genetic data](#) that is being amassed daily. NCI's goal for the GDC is to be a resource for all investigators to generate hypotheses and make new discoveries from the data."

The GDC builds upon the Bionimbus Protected Data Cloud, a pilot cloud-based system developed by Grossman that was the first to be approved by the National Institutes of Health to hold cancer genomic data from projects such as TCGA.

The GDC will be built over a number of years to ensure individual projects can be combined to create broadly useful and accessible datasets and to inform guidelines for social, ethical, and legal issues that could arise as datasets become widely shared.

The GDC will be constructed and operated with NCI funding through a subcontract from Leidos Biomedical Research, Inc. at the Frederick National Laboratory for Cancer Research. The Ontario Institute of Cancer Research is developing some of the components of the GDC through a subcontract with the University of Chicago.

Provided by University of Chicago Medical Center  
APA citation: University of Chicago to establish Genomic Data Commons (2014, December 2) retrieved 12 September 2022 from <https://medicalxpress.com/news/2014-12-university-chicago-genomic-commons.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.*