

Cancer Genomics Hub adds childhood cancer data

January 9 2014, by Tim Stephens

Researchers studying the genetics of childhood cancers now have access to a large and growing set of genomic data through the [Cancer Genomics Hub](#) (CGHub) operated by the University of California, Santa Cruz. The data come from a National Cancer Institute (NCI) initiative called TARGET (Therapeutically Applicable Research to Generate Effective Treatments), which aims to determine the molecular changes that drive the development and progression of five major types of childhood cancer.

A UCSC team led by bioinformatics expert David Haussler established CGHub in 2012 to manage primary sequence data from all of NCI's [cancer](#) genomics research programs. These programs, which aim to improve cancer treatment by identifying the genetic drivers of cancer, require massive amounts of data to be shared among researchers throughout the country. Tissue samples from thousands of cancer patients are sent to sequencing centers for analysis, yielding huge files that can amount to more than 400 gigabytes (GB) per patient. The sequencing centers upload the files to CGHub, where the data are validated, organized, stored, and made available for downloading by researchers.

Cancer researchers who use CGHub are enthusiastic about its performance, said Haussler, a distinguished professor of biomolecular engineering at UCSC's Baskin School of Engineering. "We really are providing larger, more extensive data sets more efficiently than ever before," he said. "I'm extremely proud of the team, and it's inspiring to

all of us to help with this effort, especially now with the addition of children's cancer data."

In recent months, downloads from CGHub have exceeded 1,000 terabytes (1 million GB) per month. Gad Getz, who leads cancer genome analysis at the Broad Institute of Harvard and MIT, told Haussler he has been able to download data from CGHub at an impressive rate of 4 terabytes per hour.

Easy Access

CGHub technical director Mark Diekhans said it is important for researchers to have easy access to these large datasets. "Accessing the data should not be an impediment to research. Making it easy requires not only the right infrastructure and good software, but also good user support," he said.

CGHub is a secure repository built with a planned storage capacity of 5 petabytes (5,000 terabytes). It uses GeneTorrent software, developed by Annai Systems, to enable very fast transfers of terabyte-scale data. CGHub staff built a specialized "data browser" that enables researchers to easily find and download the sequence files they need.

The first data uploaded to CGHub came from The Cancer Genome Atlas (TCGA) program, which is characterizing over 25 major types and subtypes of adult cancer. Led by NCI and the National Human Genome Research Institute, TCGA so far has produced nearly 500 terabytes of [genomic data](#) that researchers can now access through CGHub. In the past year, TCGA has yielded a series of landmark publications giving scientists a new understanding of several major types of cancer, including breast cancer, ovarian cancer, colorectal cancer, and, more recently, endometrial cancer and acute myeloid leukemia. These studies have revealed common genetic changes shared by different cancers and

suggest new ways to classify cancers into subtypes that might be targeted by different treatments.

The TARGET program promises to deliver similar insights into children's cancers. About 400 terabytes of TARGET data have been made available on CGHub.

According to Haussler, the genetic changes in children's cancers tend to be less complex than those in adult cancer, which may make it easier to understand how those changes lead to cancer. "We may be able to decode children's cancers sooner than adult cancers, but we still need an enormous amount of data to figure this out. Our goal with CGHub is to have all this data at the ready for researchers," he said.

Identifying mutations

An important part of CGHub's work involves ensuring the compatibility of data coming from different sequencing centers. Different centers have used different methods to identify mutations in the genome sequences, so CGHub scientists have designed benchmarks and validation exercises to improve the agreement between different methodologies.

"We've done a lot of work to improve the precision of the diagnosis of mutations in tumors," Haussler said. "This is an essential prelude to what many feel will become common practice in cancer medicine. Knowing where the mutations are and how they operate is a key to having precision medicine for cancer. If we can build these tools and demonstrate their usefulness, we hope that someday it will be a routine part of cancer care to have your tumor sequenced so your doctor knows exactly what the mutations are and what treatments will be most effective."

In addition to TCGA and TARGET, data from the Cancer Genome Characterization Initiative (CGCI), which focuses on various types of pediatric and adult cancers, will be available from CGHub beginning mid-2014. Access to all of these datasets is restricted to researchers approved by the National Institutes of Health.

There is, however, one dataset on CGHub that is available for public access. This is the Cancer Cell Line Encyclopedia (CCLE), which provides genomic data for about 1,000 cell lines that are widely used in laboratory research. The CCLE project is a collaboration between the Broad Institute and the Novartis Institutes for Biomedical Research, partially funded by the NCI via TCGA.

"These genomes come from cell lines that are no longer associated with a particular person, so there aren't privacy concerns about making them publicly accessible," said Linda Rosewood, CGHub program director. "This [data](#) set is very important as a research tool, but it can also be a valuable teaching tool that students can use to learn about bioinformatics and cancer genomics."

Provided by University of California - Santa Cruz

Citation: Cancer Genomics Hub adds childhood cancer data (2014, January 9) retrieved 19 April 2024 from <https://medicalxpress.com/news/2014-01-cancer-genomics-hub-childhood.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>
--