

Large collection of brain cancer data now easily, freely accessible to global researchers

August 14 2018



A valuable cache of brain cancer biomedical data has been made freely available to researchers worldwide, say researchers at Georgetown Lombardi Comprehensive Cancer Center. The dataset, REMBRANDT (REpository for Molecular BRAin Neoplasia DaTa) hosted and supported by Georgetown, is one of only two such large collections in

the country.

Information about the brain [cancer data](#) collection, which contains information on 671 adult patients collected from 14 contributing institutions, is detailed in [Scientific Data](#), an open-access journal (Nature). Already, thousands of researchers in the U.S. and internationally log on to the data site on a daily basis, and word about the resource is expected to increase its use, says Subha Madhavan, Ph.D., chief data scientist at Georgetown University Medical Center and director of the Innovation Center for Biomedical Informatics (ICBI) at Georgetown Lombardi.

The Georgetown data resource is unique in several ways. One is that it contains genomic information, collected from volunteer patients who allowed their tumors to be sampled, as well as diagnostic (including brain scans), treatment and outcomes data. Most collections contain either one or the other.

Additionally, the [data collection](#) interface is extraordinarily easy to use, Madhavan says.

"It sits on Amazon Web Services, and has a simple web interface access to data and analysis tools. All a researcher needs is a computer and an internet connection to log onto this interface to select, filter, analyze and visualize the brain tumor datasets.

"We want this data to be widely used by the broadest audience—the entire biomedical research community—so that imagination and discovery is maximized," says first author on the paper Yuriy Gusev, Ph.D., associate professor and a faculty member of the ICBI. "Our common goal is to tease apart the clues hidden within this biomedical and clinical information in order to find ways that advance diagnostic and clinical outcomes for these patients."

"We are just beginning to understand the science of how these cancers evolve and how best to treat them, and datasets like this will likely be very helpful," Madhavan says.

The REMBRANDT dataset was originally created at the National Cancer Institute and funded by Glioma Molecular Diagnostic Initiative led by co-authors Howard Fine, MD, from New York Presbyterian Hospital, and Jean-Claude Zenklusen, Ph.D., from the National Cancer Institute. They collected the data from 2004-2006.

The NCI transferred the data to Georgetown in 2015, and it is now physically located on the Georgetown Database of Cancer (G-DOC), a cancer data integration and sharing platform for hosting alongside other cancer studies. G-DOC investigators, led by Madhavan, developed novel analytical tools to process the information anew.

The genomic data includes the specific [genes](#) within individual tumors that are either over-expressed or under-expressed as well as the number of times that gene is repeated within a chromosome.

"We inherit two copies of a gene—one from Mom and one from Dad—but in cancer cells, DNA segments containing important tumor suppressor or onco- genes can be entirely deleted or amplified. It isn't unusual to see a chromosome within a tumor that has 11 copies of a gene, each of which may be producing a toxic protein that helps the cancer grow uncontrollably," she says.

The data collection also includes information on RNA, which is produced by genes (DNA) and can be measured to assess genes that are dysregulated.

Researchers can search their gene of interest, check their expression and amplification status and link that to clinical outcomes, Madhavan says.

They can save their findings to their workspace on the G-DOC site and share with their collaborators. Given the approximately 20,000 protein coding genes in the human genome, and the variety of brain cancer tumor types, "it will take a big village—really a vast metro area—of investigators to understand the bases of these tumors and to effectively develop treatments that target them."

REMBRANDT includes [genomic data](#) from 261 samples of glioblastoma, 170 of astrocytoma, 86 tissues of oligodendroglioma, and a number that are mixed or of an unknown subclass. Outcomes data include more than 13,000 data points.

Provided by Georgetown University Medical Center

Citation: Large collection of brain cancer data now easily, freely accessible to global researchers (2018, August 14) retrieved 24 April 2024 from <https://medicalxpress.com/news/2018-08-large-brain-cancer-easily-freely.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>
--