

## Access the NCI Cloud Resources

The NCI Cloud Resources are designed to provide access to innovative methods for accessing and analyzing large collections of genomic data. The goals of the Cloud Resources are to eliminate the need to download large amounts of NCI-generated genomic and related data, and provide cost-effective compute resources to the **cancer research community**. The Resources bring data and analysis tools together on a cloud platform with an Application Programming Interface (API) that provides secure access. The **Broad Institute**, the **Institute for Systems Biology**, and **Seven Bridges**, have each developed unique infrastructure with a variety of tools to access, explore, and analyze molecular data. The NCI encourages the research community to use the Cloud Resources for research and analysis, and to contribute data and tools for collaboration. The Cloud Resources allow users to run best practice tools and pipelines already implemented, or upload their own data and/or analysis methods to workspaces.

### System Authentication & Data Authorization

To use the Cloud Resources, users need to register and establish an account with each. Open-access data from The Cancer Genome Atlas (TCGA) is available on all three platforms after users agree to the data use restrictions and requirements outlined in the [TCGA publication guidelines](#). Open access data from The Cancer Imaging Archive (TCIA) and from the Clinical Proteomic Tumor Analysis Consortium (CPTAC) will also be made available.

Researchers requiring data from the Controlled Access tier (Controlled Data) for their studies need to obtain an approved Data Access Request through the [Database of Genotypes and Phenotypes \(dbGaP\)](#) and agree with all [TCGA Data Use Certifications](#) as well as the TCGA publication guidelines. All account holders must have such access, or be designated downloaders, in order to access controlled access data; help to assign downloaders to dbGaP data is available in this [video](#). To register for dbGaP access, users must have either an [eRA Commons](#) or NIH user account.

It is important to keep in mind that TCGA data hosted on the GDC and on the Cloud Resources may not be completely synchronized. This is because the timing of downloads by each of the platforms, and because the GDC hosts a broader set of data than the Cloud Resources (e.g., archived data). This issue will be addressed in the future, as the Cloud Resources switch from hosting their own set of data to accessing the data maintained by the GDC in a commercial cloud.

- **Broad Institute FireCloud:**

FireCloud is modeled after the Broad Institute's [Firehose](#) analysis infrastructure and includes tutorial workspaces as well as curated open and controlled-access TCGA workspaces that users can clone. FireCloud currently runs on the Google Cloud Platform. In order to create or clone a workspace, users must establish a Google billing account or request access to an existing project. A new Google billing account gives users \$300 of credits to get started. [Register](#) for an account, learn more about [projects and billing](#), and get support through the [help forum](#).

- **Institute for Systems Biology (ISB) Cancer Genomics Cloud:**

The ISB platform leverages many aspects of the Google Cloud Platform (GCP), including Cloud Storage, Compute Engine, BigQuery, and Genomics. Supported use-cases include computationally intensive analysis of sequence data, and integrative analyses of molecular and clinical data using R, Python, and/or SQL. Sign-in with Google credentials to access the [WebApp](#). Request a GCP project or link an existing GCP project by writing to [request-gcp@isb-cgc.org](mailto:request-gcp@isb-cgc.org). Get [help](#) with account set up and data access.

- **Seven Bridges (SB) Cancer Genomics Cloud:**

The [Seven Bridges Cloud](#) enables researchers to collaborate on the analysis of large cancer genomics datasets in a secure, reproducible and scalable manner. Deployed on Amazon Web Services, the SB Cloud has a rich query system that allows researchers to find exactly the data they are interested in and combine it with their own private data. Native implementation of the [Common Workflow Language](#) specification makes it easy for developers, analysts and bench biologists to deploy, customize and run reproducible analysis methods to learn from genomics data faster. [Register](#) for an account, [learn how to access TCGA data](#) or [follow a tutorial](#) to get started.

## Glossary of Links

### dbGap

- Database of Genotypes and Phenotypes (dbGaP): <http://www.ncbi.nlm.nih.gov/gap>
- Video instructions for assigning authorized dbGaP downloaders: <http://bit.ly/24JX6xO>

### eRA Commons

- eRA Commons User Account Login: <https://public.era.nih.gov/commons/public/login.do>

### TCGA

- TCGA Data Use Certifications: [http://cancergenome.nih.gov/pdfs/Data\\_Use\\_Certv082014](http://cancergenome.nih.gov/pdfs/Data_Use_Certv082014)
- TCGA Publication Guidelines: <http://cancergenome.nih.gov/publications/publicationguidelines>

## NCI Cloud Resources Website

- [cbiit.nci.nih.gov/cloudresources](http://cbiit.nci.nih.gov/cloudresources)

### Broad Institute

- FireCloud – Broad Institute Cancer Genomics Cloud: <http://firecloud.org>
- Register: <https://portal.firecloud.org/>
- Projects and Billing: <https://goo.gl/u6cYvt>
- Firehose – Broad Institute Analytics Infrastructure: <https://www.broadinstitute.org/cancer/cga/Firehose>
- Help Forum: <http://gatkforums.broadinstitute.org/firecloud>

### Institute for Systems Biology

- ISB Cancer Genomics Cloud: [www.isb-cgc.org](http://www.isb-cgc.org)
- WebApp Sign-in (using Google credentials): <https://isb-cgc.appspot.com/>
- Online Help: <http://isb-cancer-genomics-cloud.readthedocs.io/en/latest/>

### Seven Bridges

- SB Cancer Genomics Cloud: <http://www.cancergenomicscloud.org>
- Account Registration: <https://cgc.sbgenomics.com/register/>
- Online Help: <http://docs.cancergenomicscloud.org/docs/>