Access the NCI Cancer Genomics Cloud Pilots

The Cancer Genomics Cloud (CGC) Pilots are designed to explore innovative methods for accessing and analyzing large collections of genomic data. The goals of the CGC Pilots 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 Pilots 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 Genomics, have each developed unique infrastructure with a variety of tools to access, explore, and analyze molecular data. The NCI encourages the research community to test the usefulness of the CGC Pilots to their own research. The CGC Pilots 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

There is no cost to use the CGC Pilot platforms; users just 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. 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. To use the CGC Pilots, users are required to include a “Cloud Use Statement” with a new dbGaP application.

- **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. Get help with account set up and data access.

- **Seven Bridges (SB) Genomics Cancer Genomics Cloud:**
  
  The Seven Bridges Cloud Pilot enables researchers to collaborate on the analysis of large cancer genomics datasets in a secure, reproducible and scalable manner. A rich query system 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. Deployed on Amazon Web Services, researchers can access up to $1,500 in compute and storage credits or request up to $10,000 to support collaborative projects. Find out more about credits, Register for an account, learn how to access TCGA data or follow a tutorial to get started.
Glossary of Links

**dbGap**

**eRA Commons**

**TCGA**

**NCI CGC Pilots Web Site**
- cbiit.nci.nih.gov/cloudpilots

**Broad Institute**
- FireCloud – Broad Institute Cancer Genomics Cloud: [http://firecloud.org](http://firecloud.org)
- Register: [https://portal.firecloud.org/](https://portal.firecloud.org/)
- Projects and Billing: [https://goo.gl/u6cYvt](https://goo.gl/u6cYvt)
- Firehose – Broad Institute Analytics Infrastructure: [https://www.broadinstitute.org/cancer/cga/Firehose](https://www.broadinstitute.org/cancer/cga/Firehose)

**Institute for Systems Biology**
- ISB Cancer Genomics Cloud: [www.isb-cgc.org](http://www.isb-cgc.org)

**Seven Bridges Genomics**
- SBG Cancer Genomics Cloud: [http://www.cancergenomicscloud.org](http://www.cancergenomicscloud.org)
- Account Registration: [https://cgc.sbgenomics.com/register/](https://cgc.sbgenomics.com/register/)
- Online Help: [http://docs.cancergenomicscloud.org/docs/](http://docs.cancergenomicscloud.org/docs/)

**Online Survey to Provide Feedback on the Pilots**
- [www.surveymonkey.com/r/NCIGenomicsCloudPilots](http://www.surveymonkey.com/r/NCIGenomicsCloudPilots)