

NCI Cloud Resources

Bringing data and computation together to create knowledge that accelerates cancer research and enables precision medicine.

The NCI Cloud Resources were developed as part of an NCI pilot program designed to explore innovative methods for accessing and computing on large scale cancer genomic data. Based on the success of the pilot, NCI will continue to sponsor the platforms that make up the Cloud Resources to provide continuing benefit to the cancer research community.

The Cloud Resources bring together analytics and data from The Cancer Genome Atlas (TCGA) and other sources with co-located computational capacity and an Application Programming Interface (API) that provides secure data access. In this model, applications are brought to the data, rather than bringing the data to the applications.



Data

- Access and analyze 11,000 TCGA samples without having to download data
- Search and view images from The Cancer Imaging Archive (TCIA)
- Access genomic data from Therapeutically Applicable Research to Generate Effective Treatments (TARGET)
- Access and analyze proteomic data from the Clinical Proteomic Tumor Analysis Consortium (CPTAC)
- Upload your own data for analysis
- Conduct and store your analyses in a workspace, and share your data, tools, and results with collaborators



Compute

- Perform large scale analysis using the elastic compute power of commercial cloud platforms
- Use best-of-breed pipelines and analytic tools
- Upload your own tools to analyze the data in the cloud



Security

- dbGaP-authorized users can access controlled TCGA and TARGET data
- Systems meet strict federal security guidelines



#NCICloud

For more information about the Cancer Genomics Cloud Resources, please visit: cbiit.nci.nih.gov/cloudresources

The goals of the NCI Cloud Resources are to democratize access to NCI-generated genomic and related data, and to create a cost-effective way to provide scalable computational capacity to the cancer research community.

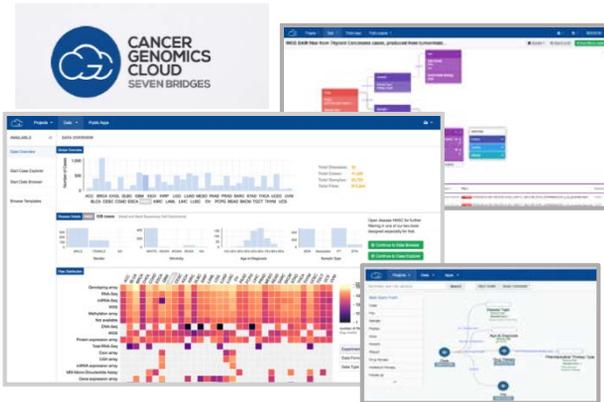
Institute for Systems Biology



www.isb-cgc.org

The **Institute for Systems Biology (ISB) Cloud** provides interactive and programmatic access to data, leveraging many aspects of the Google Cloud Platform. The **interactive ISB-CGC web-app** allows scientists to interactively define and compare cohorts, examine underlying molecular data for specific genes or pathways of interest, and share insights with collaborators. For computational users, programmatic interfaces and GCP tools such as **BigQuery, Genomics, and Compute Engine** allow users to perform complex queries from R or Python scripts, or run Dockerized workflows on sequence data available in cloud storage.

Seven Bridges



www.cancergenomicscloud.org

Seven Bridges Cancer Genomics Cloud 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 the exact data of interest 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.

Broad Institute



www.firecloud.org

Broad Institute FireCloud is modeled after their **Firehose analysis infrastructure** and facilitates collaboration and provides a robust, scalable platform accessible to the community at-large. Using the elastic compute capacity of Google Cloud, FireCloud empowers analysts, tool developers, and production managers to perform large-scale analysis, engage in data curation, and store or publish results. Users can upload their own analysis methods and data to workspaces or run the **Broad's best practice tools and pipelines** on pre-loaded data.