The Cancer Genome Atlas:
Navigating the Data Portal

http://cancergenome.nih.gov/dataportal/data/about
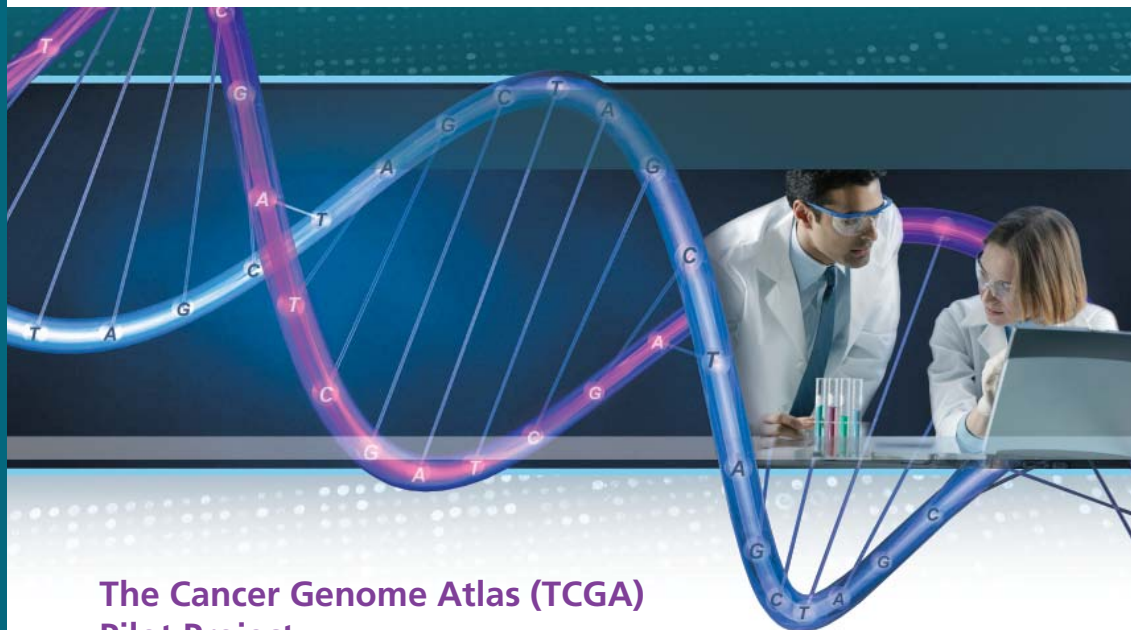
The Cancer Genome Atlas (TCGA) Pilot Project

The Cancer Genome Atlas, a pilot project of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), is a large-scale collaborative effort to understand the molecular basis of cancer through the application of genome analysis technologies, including but not limited to identifying mutations in DNA sequence, copy number variation and alterations in methylation status. The pilot project is focused on three types of cancers: glioblastoma multiforme (GBM), squamous carcinoma of the lung, and serous cystadenocarcinoma of the ovary.
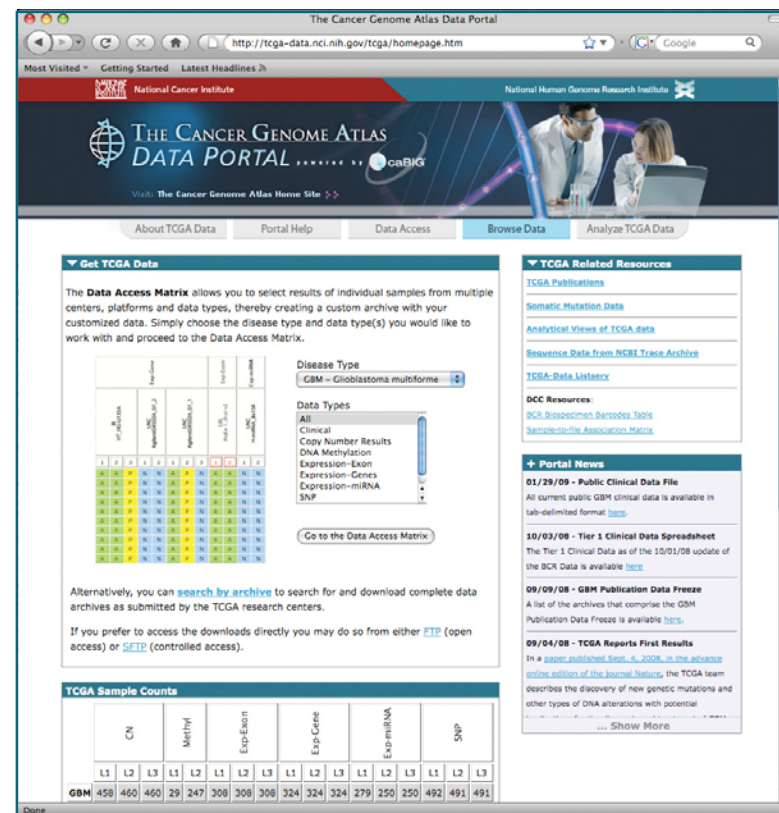
TCGA is unique in the manner in which it collaborates with a broad cross-section of the cancer research community, including basic and clinical researchers, oncologists, nurses, bioethicists, cancer patients, and advocates. Approximately two dozen public and private institutions in the United States and across the globe are involved in the TCGA Research Network. The network is actively engaged in data generation, characterization and analysis and is enabling integrated exploration of these cancer genome data sets by the entire cancer research community. Through this coordinated and collaborative effort, TCGA is providing insights that will accelerate the development of next-generation diagnostic biomarkers and targeted therapies. The discoveries made as a result of TCGA will lay the foundation for a more personalized approach to cancer prevention, diagnostics, and treatment.

To read more about TCGA, its goals, and its discoveries, please visit: http://cancergenome.nih.gov/

TCGA Data Portal – What is It?

TCGA Data Portal is a Web-based platform for cancer researchers to search, download, and analyze data sets generated by TCGA. The portal contains all TCGA data sets pertaining to clinical information, genomic characterization, and high-throughput sequencing analysis of the tumor genomes.

Data currently available through the portal include molecular characterization data sets for ovarian cancer, and molecular characterization, high-throughput sequencing, and clinical data for glioblastoma multiforme (GBM) tumors. New data are derived on an ongoing basis from TCGA analyses and are deposited into databases for download in the portal.

To download the Data Request Form and Data Use Certification, please visit:

http://tcga.cancer.gov/dataportal/data/access/closed

http://tcga-data.nci.nih.gov/tcga/homepage.htm

With large volumes of genomic information derived from human tumor specimens collected from patient populations, the aggregated data generated from TCGA are available in two tiers: **Open-Access Data tier** – a publicly accessible tier of data that cannot be used to identify a specific individual; and **Controlled-Access Data tier** – a controlled, password-protected tier containing information such as clinical data that could be used to identify patient participants. To access the protected data, a user must complete a Data Access Request (DAR) form and Data Use Certification (DUC); and access is granted on a case-by-case basis.

Researchers are able to apply the data available from TCGA Data Portal to further clinical advances in cancer prevention, diagnosis, and treatment.

How does the Data Portal Work?

Providing access to advanced analysis tools, the portal allows the user to select individual samples from multiple centers, platforms, and data types to create a customized data set. In addition, users can search and download complete data archives as submitted by TCGA research centers for clinical, characterization, and sequencing data of the tumor genomes.

How to Access and Analyze TCGA Data:

Access TCGA Data Portal:

http://cancergenome.nih.gov/dataportal/data/about

Analyze TCGA Data:

http://tcga-data.nci.nih.gov/docs/related/analytical_views.html

Tools Featured in the TCGA Data Portal

- **Cancer Molecular Analysis (CMA) Portal:** A component of the National Cancer Institute's caBIG® initiative, the CMA enables researchers to integrate, visualize, and explore genomic characterization data from translational studies, including TCGA.

- **Cancer Genome Workbench (CGWB):** A computational platform, CGWB integrates tumor mutation profiles with multidimensional molecular and clinical data.

- **Integrative Genomics Viewer (IGV):** A fast, flexible visualization tool for genomic data that integrates datasets from various platforms and sources, including genetic variation, gene/microRNA expression, epigenetic data, RNAi screens, genomic features, and annotation of samples.

- **MSKCC CancerGenomics Analysis:** This tool provides detailed analysis results and analysis tools in the areas of signaling pathways, mutations, and copy number alterations.

The National Cancer Institute continues to support the development of novel tools for Data Portal users to assist in overcoming the limitations of current methods for analyzing large, complex data sets.

Responsible Data Use

In promoting the responsible dissemination of TCGA dataset analyses, TCGA follows a stringent informed consent and patient protection process. TCGA approved data users are strongly encouraged to visit the data use policy and publications guidelines for information and guidance on how to use TCGA data in publications.

Learn more about TCGA patient protection policies and data use guidelines at:

http://cancergenome.nih.gov/policies/index.asp

How Will TCGA Benefit Cancer Research and Patients?

Data from the TCGA Pilot Project provide researchers and clinicians with an early glimpse of what promises to become an unprecedented, comprehensive "atlas" of molecular information describing genomic changes associated with specific cancers. TCGA will ultimately enable researchers to analyze and employ the data in their own research to advance the molecular understanding of cancer and pave the way for a more targeted, personalized approach to cancer prevention and treatment.

TCGA data are already generating promising research findings with the publication of its first large-scale, comprehensive study on glioblastoma multiforme (GBM). TCGA identified many gene mutations involved in GBM, including three previously unrecognized mutations that occur with significant frequency and the delineation of core pathways disrupted in this type of brain cancer. Among the most exciting results is an unexpected observation that points to a potential mechanism of resistance to a standard chemotherapy drug, temozolomide, used to treat glioblastoma patients. These findings demonstrate TCGA's promise in generating further discoveries as users integrate more and more large-scale data sets generated by TCGA.

Find Out More About TCGA

For more information about The Cancer Genome Atlas Pilot Project, please contact:

The Cancer Genome Atlas Project

National Cancer Institute

National Institutes of Health

Building 31, Room 10A07

31 Center Drive, MSC 2580

Bethesda, MD 20892-2580

301-451-8027

For additional information about the sponsoring institutes, please visit the following websites:

National Cancer Institute

http://cancer.gov

National Human Genome Research Institute

http://genome.gov