The Cancer Genome Atlas Glioblastoma Multiforme Collection (TCGA-GBM)

Reproduced from https://wiki.cancerimagingarchive.net

Summary

The Cancer Genome Atlas Glioblastoma Multiforme (TCGA-GBM) data collection is part of a larger effort to build a research community focused on connecting cancer phenotypes to genotypes by providing clinical images matched to subjects from The Cancer Genome Atlas (TCGA). Clinical, genetic, and pathological data resides in the Genomic Data Commons (GDC) Data Portal while the radiological data is stored on The Cancer Imaging Archive (TCIA).

Matched TCGA patient identifiers allow researchers to explore the TCGA/TCIA databases for correlations between tissue genotype, radiological phenotype and patient outcomes. Tissues for TCGA were collected from many sites all over the world in order to reach their accrual targets, usually around 500 specimens per cancer type. For this reason the image data sets are also extremely heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. In most cases the images were acquired as part of routine care and not as part of a controlled research study or clinical trial.

CIP TCGA Radiology Initiative

Imaging Source Site (ISS) Groups are being populated and governed by participants from institutions that have provided imaging data to the archive for a given cancer type. Modeled after TCGA analysis groups, ISS groups are given the opportunity to publish a marker paper for a given cancer type per the guidelines in the table above. This opportunity will generate increased participation in building these multi-institutional data sets as they become an open community resource. Learn more about the TCGA Glioma Phenotype Research Group.

Acknowledgements

We would like to acknowledge the individuals and institutions that have provided data for this collection:

- Henry Ford Hospital, Detroit, MI - Special thanks to **Lisa Scarpace** and **Tom Mikkelsen, MD** from the Department of Neurosurgery, Hermelin Brain Tumor Center.
- University of California, San Francisco, CA - Special thanks to **Soonmee Cha, MD** from the Department of Neurological Surgery, Brain Tumor Research Center.
- MD Anderson Cancer Center, Houston, TX - Special thanks to **Sujaya Rao** and **Sangeeta Tekchandani** from the Office of Translational/Clinical Research.
- Emory University, Atlanta, GA - Special thanks to **David Gutman, MD, Ph.D.** and **Joel Saltz, MD, Ph.D.** from the Center for Comprehensive Informatics.
- Thomas Jefferson University, Philadelpha, PA - Special thanks to **Nancy Pedano** and **Adam E. Flanders, MD** fr om the Department of Radiology, Jefferson Medical College.
- CWRU School of Medicine, Cleveland, OH - Special thanks to **Jill Barnholtz-Sloan, Ph.D.** and **Quinn Ostrom, MA, MPH** from Case Comprehensive Cancer Center.
- Duke University School of Medicine, Durham, NC - Special thanks to **Daniel Barboriak, MD** and Laura J Pierce.
- Fondazione IRCCS Instituto Neuroligico C. Besta, Milan, Italy - Special thanks to **Domenico Aquino** and **Alessa ndro Perin MD**.

Data Access

Data Access

Some data in this collection contains images that could potentially be used to reconstruct a human face. To safeguard the privacy of participants, users must sign and submit a TCIA Restricted License Agreement to help@cancerimagingarchive .net before accessing the data.

Reproduced from https://wiki.cancerimagingarchive.net

Data Type	Download all or Query/Filter	License
Images (DICOM, 73.5GB)	*Download Search* (Download requires the NBIA Data Retriever)	TCIA Restricted

Click the Versions tab for more info about data releases.

Additional Resources for this Dataset

The NCI Cancer Research Data Commons (CRDC) provides access to additional data and a cloud-based data science infrastructure that connects data sets with analytics tools to allow users to share, integrate, analyze, and visualize cancer research data.

- Imaging Data Commons (IDC) (Imaging Data)
- Genomic Data Commons (GDC) (Genomic, Digitized Histopathology & Clinical Data)

Third Party Analyses of this Dataset

TCIA encourages the community to publish your analyses of our datasets. Below is a list of such third party analyses published using this Collection:

- MRQy quality measures for TCIA MRI datasets (MRQy-Quality-Measures)
- Segmentation Labels and Radiomic Features for the Pre-operative Scans of the TCGA-GBM collection (BraTS-TCGA-GBM)
- DICOM-SEG Conversions for TCGA-LGG and TCGA-GBM Segmentation Datasets (DICOM-Glioma-SEG)
- MR Imaging Predictors of Molecular Profile and Survival: Multi-institutional Study of the TCGA Glioblastoma Data Set (TCGA-GBM-Radiogenomics)
- Outcome Prediction in Patients with Glioblastoma by Using Imaging, Clinical, and Genomic Biomarkers: Focus on the Nonenhancing Component of the Tumor (GBM-MR-NER-Outcomes)
- Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features (TCGA-GBM-QI-Radiogenomics)
- Image Data Used in the Simulations of "The Role of Image Compression Standards in Medical Imaging: Current Status and Future Trends"
- Glioblastoma: Imaging Genomic Mapping Reveals Sex-specific Oncogenic Associations of Cell Death (Glioblastoma-Genomic-Mapping)
- Spatial Habitat Features derived from Multiparametric Magnetic Resonance Imaging data from Glioblastoma Multiforme cases (Spatial-Features-MRI-GBM)

Detailed Description

Detailed Description

Image Statistics	Radiology
Modalities	CT,MR,DX
Number of Participants	262
Number of Studies	575

Reproduced from https://wiki.cancerimagingarchive.net

Number of Series	5,412
Number of Images	481,158
Images Size (GB)	73.5

GDC Data Portal - Clinical and Genomic Data

The GDC Data Portal has extensive clinical and genomic data, which can be matched to the patient identifiers of the images here in TCIA. Below is a snapshot of clinical data extracted on 1/5/2016:

- TCGA-GBM Clinical Data.zip (**NOTE:** this is just a representative sample of what's available. Visit the GDC Data Portal to obtain the latest/complete data set)

Explanations of the clinical data can be found on the Biospecimen Core Resource Clinical Data Forms linked below:

- GBM Case Quality Control Form
- GBM Enrollment Form
- GBM Follow-Up Form

A Note about TCIA and TCGA Subject Identifiers and Dates

Subject Identifiers: a subject with radiology images stored in TCIA is identified with a Patient ID that is identical to the Patient ID of the same subject with demographic, clinical, pathological, and/or genomic data stored in TCGA. For each TCGA case, the baseline TCGA imaging studies found on TCIA are pre-surgical.

Dates: TCIA and TCGA handle dates differently, and there are no immediate plans to reconcile:

- TCIA Dates: dates (be they birth dates, imaging study dates, etc.) in the Digital Imaging and Communications in Medicine (DICOM) headers of TCIA radiology images have been offset by a random number of days. The offset is a number of days between 3 and 10 years prior to the real date that is consistent for each TCIA image-submitting site and collection, but that varies among sites and among collections from the same site. Thus, the number of days between a subject's longitudinal imaging studies are accurately preserved when more than one study has been archived while still meeting HIPAA requirements.
- TCGA Dates: the patient demographic and clinical event dates are all the number of days from the index date, which is the actual date of pathologic diagnosis. So all the dates in the data are relative negative or positive integers, except for the "days_to_pathologic_diagnosis" value, which is 0 – the index date. The years of birth and diagnosis are maintained in the distributed clinical data file. The NCI retains a copy of the data with complete dates, but those data are not made available.With regard to other TCGA dates, if a date comes from a HIPAA "covered entity's" medical record, it is turned into the relative day count from the index date. Dates like the date TCGA received the specimen or when the TCGA case report form was filled out are not such covered dates, and they will appear as real dates (month, day, and year).

Citations & Data Usage Policy

Citations & Data Usage Policy

Users must abide by the TCIA Data Usage Policy and Restrictions. Attribution should include references to the following citations:

Reproduced from https://wiki.cancerimagingarchive.net

Data Citation

Scarpace, L., Mikkelsen, T., Cha, S., Rao, S., Tekchandani, S., Gutman, D., Saltz, J. H., Erickson, B. J., Pedano, N., Flanders, A. E., Barnholtz-Sloan, J., Ostrom, Q., Barboriak, D., & Pierce, L. J. (2016). **The Cancer Genome Atlas Glioblastoma Multiforme Collection (TCGA-GBM) (Version 5) [Data set].** The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.RNYFUYE9

Acknowledgement

"The results <published or shown> here are in whole or part based upon data generated by the TCGA Research Network: http://cancergenome.nih.gov/."

TCIA Citation

Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., & Prior, F. (2013). The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. In Journal of Digital Imaging (Vol. 26, Issue 6, pp. 1045–1057). Springer Science and Business Media LLC. https://doi.org/10.1007/s10278-013-9622-7

Other Publications Using This Data

TCIA maintains a list of publications which leverage our data. If you have a publication you'd like to add please contact TCIA's Helpdesk.

Versions

Version 5 (Current): Updated 2023/08/07

Repaired a DICOM tag(0008,0005) to value "ISO_IR 100" in 5 series of ID TCGA-08-0380.

Version 4: Updated 2020/05/29

Data Type	Download all or Query/Filter
Images (DICOM, 73.5GB)	*Download* *Search* (Download requires the NBIA Data Retriever)
Tissue Slide Images (web)	*Search*
Clinical Data (TXT)	*Download*
Biomedical Data (TXT)	*Download*
Genomics (web)	*Search*

Reproduced from https://wiki.cancerimagingarchive.net

Updated clinical data link with latest spreadsheets from GDC. Added new biomedical spreadsheets from GDC.

Version 3: Updated 2015/09/16

Data Type	Download all or Query/Filter
Images (DICOM, 73.5GB)	Download Search (Download requires the NBIA Data Retriever)
Clinical Data (TXT)	Download
Genomics (web)	Search

1 new subject added.

Version 2: Updated 2016/01/05

Data Type	Download all or Query/Filter
Images (DICOM, 72.8GB)	Download (Download requires the NBIA Data Retriever)
Clinical Data (TXT)	Download
Genomics (web)	Search

Extracted latest release of clinical data (TXT) from the GDC Data Portal.

Version 1: Updated 2014/12/30

Data Type	Download all or Query/Filter
Images (DICOM, 72.8GB)	Download (Download requires the NBIA Data Retriever)
Clinical Data (TXT)	Download
Genomics (web)	Search

Reproduced from https://wiki.cancerimagingarchive.net

On 03-01-2013 available subjects in TCIA were reduced from 279 to 240 because some cases were subsequently excluded from the GDC Data Portal.

Reproduced from https://wiki.cancerimagingarchive.net