The Cancer Genome Atlas Sarcoma Collection (TCGA-SARC)

Reproduced from https://wiki.cancerimagingarchive.net

Summary

The Cancer Genome Atlas Sarcoma (TCGA-SARC) 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 CIP TCGA Radiology Initiative.

Acknowledgements

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

- Roswell Park Cancer Institute, Buffalo NY - Special thanks to **Charles Roche, MD; Ermalinda Bonaccio, MD;** and **Joe Filippini.**

Data Access

Data Access

Data Type	Download all or Query/Filter	License
Images (DICOM, 2.8GB)	*Download* *Search* (Download requires the NBIA Data Retriever)	CC BY 3.0

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)

Detailed Description

Reproduced from https://wiki.cancerimagingarchive.net

Detailed Description

Image Statistics	
Modalities	CT, MR
Number of Participants	5
Number of Studies	6
Number of Series	33
Number of Images	5,653
Images Size (GB)	2.8

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 on the images here in TCIA. Below is a snapshot of clinical data extracted on 1/5/2016.

- TCGA-SARC Clinical Data.zip

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

- Sarcoma Case Quality Control Form
- Sarcoma Enrollment Form
- Sarcoma 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).

Reproduced from https://wiki.cancerimagingarchive.net

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:

Data Citation

Roche, C., Bonaccio, E., & Filippini, J. (2016). **The Cancer Genome Atlas Sarcoma Collection (TCGA-SARC) (Version 3) [Data set].** The Cancer Imaging Archive. https://doi.org/10.7937/K9/TCIA.2016.CX6YLSUX

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. At this time we are not aware of any manuscripts based on this data. If you have a manuscript you'd like to add please contact TCIA's Helpdesk.

Versions

Version 3 (Current): Updated 2020/05/29

Data Type	Download all or Query/Filter
Images (DICOM, 2.8GB)	*Download* *Search* (Download requires the NBIA Data Retriever)
Tissue Slide Images (web)	*Search*
Clinical Data (TXT)	*Download*
Biomedical Data (TXT)	*Download*
Genomics (web)	*Search*

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

Version 2: Updated 2016/01/05

Data Type	Download all or Query/Filter

Images (DICOM, 2.8GB)	*Download* (Download requires the NBIA Data Retriever)
Clinical Data (TXT)	*Download*
Genomics (web)	*Search*

Added a data dump from the GDC Data Portal clinical data.

Version 1: Updated 2014/12/30

Data Type	Download all or Query/Filter
Images (DICOM, 2.8GB)	*Download* (Download requires the NBIA Data Retriever)
Genomics (web)	*Search*

Reproduced from https://wiki.cancerimagingarchive.net