TCGA-LUAD

Summary

The Cancer Genome Atlas Lung Adenocarcinoma (TCGA-LUAD) 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 Lung Phenotype Research Group.

Acknowledgements

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

- Washington University in St. Louis, St. Louis, MO - Special thanks to **Brad Albertina** from the Barnes-Jewish Hospital Department of Radiology and **Mark Watson, MD, Ph.D.** from the Tissue Procurement and Multiplexed Gene Analysis Laboratories, Washington University School of Medicine.
- University of Pittsburgh/UPMC, Pittsburgh, PA - Special thanks to **Chandra Holback, MD** and **Rose Jarosz**.
- University of North Carolina, Chapel Hill, NC - Special thanks to **J. Keith Smith**, **MD, Ph.D.** and **Shanah Kirk** from the Department of Radiology.
- Lahey Hospital & Medical Center, Burlington, MA - Special thanks to **John Lemmerman, RT** and **Kimberly Reiger-Christ, PhD**, Cancer Research, Sophia Gordon Cancer Center.

Data Access

Data Access

Choosing the Download option will provide you with a file to launch the TCIA Download Manager to download the entire collection. If you want to browse or

filter the data to select only specific scans/studies please use the Search By Collection option.

Data Type	Download all or Query/Filter
Images (DICOM, 18.3GB)	Download Search (Download requires the NBIA Data Retriever.)
Tissue Slide Images (web)	Search
Clinical Data (TXT)	Download
Genomics (web)	Search

Click the Versions tab for more info about data releases.

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:

- Crowds Cure Cancer: Data collected at the RSNA 2017 annual meeting
- Tumor-Infiltrating Lymphocytes Maps from TCGA H&E Whole Slide Pathology Images

Detailed Description

Detailed Description

Image Statistics	
Modalities	CT, PT, NM
Number of Patients	69
Number of Studies	152
Number of Series	624
Number of Images	48,931
Images Size (GB)	18.3

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-LUAD Clinical Data.zip

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

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

A Note about TCIA and TCGA Subject Identifiers and Dates

NOTE: On 12/3/12 TCIA staff were alerted to the fact that the middle 2 digits in the patient identifiers of images were not consistent with those on the TCGA Data Portal (e.g. patient TCGA-93-Z011 should have been TCGA-17-Z011). This was corrected and the data was re-posted on 12/5/12 so that all subjects now have 17 digits as intended.

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

Please be sure to include the following attribution and citations in your work if you use this data set:

TCGA Attribution

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

Data Citation

Albertina, B., Watson, M., Holback, C., Jarosz, R., Kirk, S., Lee, Y., … Lemmerman, J. (2016). Radiology Data from The Cancer Genome Atlas Lung Adenocarcinoma [TCGA-LUAD] collection. The Cancer Imaging Archive. http://doi.org/10.7937/K9/TCIA.2016.JGNIHEP5

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. **The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository**, Journal of Digital Imaging, Volume 26, Number 6, December, 2013, pp 1045-1057. (paper)

Other Publications Using This Data

TCIA maintains a list of publications which leverage our data.

1. Choi, Hongyoon and Kwon Joong Na. "Integrative Analysis of Imaging and Transcriptomic Data of the Immune Landscape Associated with Tumor Metabolism in Lung Adenocarcinoma: Clinical and Prognostic Implications." THERANOSTICS, vol. 8, no. 7, 2018, pp. 1956-1965, doi: 10.7150/thno.23767.
2. Dara, S et al. "Feature Extraction in Medical Images by Using Deep Learning Approach." International Journal of Pure and Applied Mathematics, vol. 120, no. 6, 2018, pp. 305-312, https://acadpubl.eu/hub/2018-120-6/1/20.pdf
3. Livieris, Ioannis et al. "Detecting Lung Abnormalities from X-Rays Using an Improved Ssl Algorithm." Electronic Notes in Theoretical Computer Science, vol. 343, 2019, pp. 19-33, doi:10.1016/j.entcs.2019.04.008.

4. Livieris, Ioannis et al. "A Weighted Voting Ensemble Self-Labeled Algorithm for the Detection of Lung Abnormalities from X-Rays." Algorithms, vol. 12, no. 3, 2019, p. 64, doi:10.3390/a12030064.
5. Matsuyama, Eri and Du-Yih Tsai. "Automated Classification of Lung Diseases in Computed Tomography Images Using a Wavelet Based Convolutional Neural Network." Journal of Biomedical Science and Engineering, vol. 11, no. 10, 2018, p. 263, doi:10.4236/jbise.2018.1110022.
6. Meldo, Anna et al. "Database Acquisition for the Lung Cancer Computer Aided Diagnostic Systems." 25th conference of FRUCT(Finnish-Russian University Cooperation in Telecommunications) Association, Nov 5-8 2019 2019. https://fruct.org/publications/fruct25/files/Mel.pdf.
7. Pathak, Yadunath et al. "An Efficient Low-Dose Ct Reconstruction Technique Using Partial Derivatives Based Guided Image Filter." Multimedia Tools and Applications, 2018, pp. 1-20, doi:10.1007/s11042-018-6840-5.
8. Singh, Apurva et al. "A Novel Imaging-Genomic Approach to Predict Outcomes of Radiation Therapy." Department of Electrical and Computer Engineering, vol. MS, George Washington University, 2019. general editor, Murray Loew.
9. Toaçar, Mesut et al. "Detection of Lung Cancer on Chest Ct Images Using Minimum Redundancy Maximum Relevance Feature Selection Method with Convolutional Neural Networks." Biocybernetics and Biomedical Engineering, 2019, doi:10.1016/j.bbe.2019.11.004.
10. Wong, Jordan et al. "Comparing Deep Learning-Based Auto-Segmentation of Organs at Risk and Clinical Target Volumes to Expert Inter-Observer Variability in Radiotherapy Planning." Radiother Oncol, vol. 144, 2019, pp. 152-158, doi:10.1016/j.radonc.2019.10.019.
11. Yu, Zexi. "Co-Segmentation Methods for Improving Tumor Target Delineation in Pet-Ct Images." Electrical and Computer Engineering, vol. Master of Science (M.Sc.), University of Saskatchewan, 2016, p. 119. general editor, Francis; Babyn Bui, Paul, (Link to Thesis)

If you have a manuscript you'd like to add please contact the TCIA Helpdesk.

Versions

Version 3 (Current): Updated 2017/01/30

Data Type	Download all or Query/Filter
Images (DICOM, 18.3GB)	Download Search (Download requires the NBIA Data Retriever.)
Clinical Data (TXT)	Download
Genomics (web)	Search

Images for 5 new subjects added.

Version 2: Updated 2016/01/05

Data Type	Download all or Query/Filter
Images (DICOM, 9.4GB)	Download Search (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 2014/03/28

Data Type	Download all or Query/Filter
Images (DICOM, 9.4GB)	Download Search (Download requires the NBIA Data Retriever.)
Clinical Data (TXT)	Download
Genomics (web)	Search