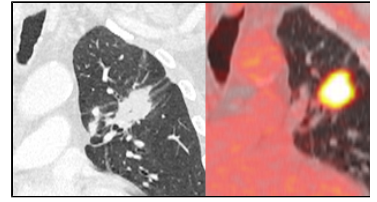


NSCLC Radiogenomics



Summary






Medical image biomarkers of cancer promise improvements in patient care through advances in precision medicine. Compared to genomic biomarkers, image biomarkers provide the advantages of being a non-invasive procedure, and characterizing a heterogeneous tumor in its entirety, as opposed to limited tissue available for biopsy. We developed a unique radiogenomic dataset from a Non-Small Cell Lung Cancer (NSCLC) cohort of 211 subjects. The dataset comprises Computed Tomography (CT), Positron Emission Tomography (PET)/CT images, semantic annotations of the tumors as observed on the medical images using a controlled vocabulary, segmentation maps of tumors in the CT scans, and quantitative values obtained from the PET/CT scans. Imaging data are also paired with gene mutation, RNA sequencing data from samples of surgically excised tumor tissue, and clinical data, including survival outcomes. This dataset was created to facilitate the discovery of the underlying relationship between genomic and medical image features, as well as the development and evaluation of prognostic medical image biomarkers.

Further details regarding this data-set may be found in Bakr, et. al, Sci Data. 2018 Oct 16;5:180202. doi: 10.1038/sdata.2018.202, <https://www.ncbi.nlm.nih.gov/pubmed/30325352>. Please direct additional inquiries to Drs. Sandy Napel (snapel@stanford.edu) or Sylvia K. Plevritis (sylvia.plevritis@stanford.edu).

Data Access

Data Access

Click the **Download** button to save a ".tcia" manifest file to your computer, which you must open with the [NBIA Data Retriever](#) . Click the **Search** button to open our Data Portal, where you can browse the data collection and/or download a subset of its contents.

Data Type	Download all or Query/Filter
Images and Segmentations (DICOM, 97.6 GB)	 
AIM Annotations (XML, zip)	
Clinical Data (csv)	
RNA sequence data (web) Note: 130 subject subset	

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 2018 annual meeting](#)
- [QIN multi-site collection of Lung CT data with Nodule Segmentations](#)
- [NSCLC Radiogenomics: Initial Stanford Study of 26 Cases](#)

Detailed Description

Detailed Description

Collection Statistics	
Modalities	CT, PT, SEG
Number of Participants	211
Number of Studies	303
Number of Series	1355
Number of Images	285,411
Image Size (GB)	97.6

This collection was originally submitted to TCIA as a 26 subject pilot data set. You can learn more about that subset of the collection in the following [Analysis Results](#) publication:

Data Citation

Napel, Sandy, & Plevritis, Sylvia K. (2014). NSCLC Radiogenomics: Initial Stanford Study of 26 Cases. The Cancer Imaging Archive. <http://doi.org/10.7937/K9/TCIA.2014.X7ONY6B1>

Citations & Data Usage Policy

Citations & Data Usage Policy

Users of this data must abide by the [Creative Commons Attribution 3.0 Unported License](#) under which it has been published. Attribution should include references to the following citations:

Data Citation

Bakr, Shaimaa; Gevaert, Olivier; Echegaray, Sebastian; Ayers, Kelsey; Zhou, Mu; Shafiq, Majid; Zheng, Hong; Zhang, Weiruo; Leung, Ann; Kadouch, Michael; Shrager, Joseph; Quon, Andrew; Rubin, Daniel; Plevritis, Sylvia; Napel, Sandy.(2017). Data for NSCLC Radiogenomics Collection. The Cancer Imaging Archive. <http://doi.org/10.7937/K9/TCIA.2017.7hs46erv>

i Publication Citation

Bakr, S., Gevaert, O., Echegaray, S., Ayers, K., Zhou, M., Shafiq, M., Zheng, H., Benson, J. A., Zhang, W., Leung, A., Kadoch, M., Hoang, C. D., Shrager, J., Quon, A., Rubin, D. L., Plevritis, S. K., & Napel, S. (2018). A radiogenomic dataset of non-small cell lung cancer. *Scientific data*, 5, 180202. <https://doi.org/10.1038/sdata.2018.202>

i Publication Citation

Gevaert, O., Xu, J., Hoang, C. D., Leung, A. N., Xu, Y., Quon, A., ... Plevritis, S. K. (2012, August). Non-Small Cell Lung Cancer: Identifying Prognostic Imaging Biomarkers by Leveraging Public Gene Expression Microarray Data—Methods and Preliminary Results. *Radiology*. Radiological Society of North America (RSNA). <http://doi.org/10.1148/radiol.12111607>

i 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

If you have a publication you'd like to add, please [contact the TCIA Helpdesk](#).

Versions




Version 3 (Current): Updated 2020/11/10

Data Type	Download all or Query/Filter
Images (DICOM, 97.6 GB)	<div style="display: flex; gap: 10px;">  Download  Search </div> <p>(Requires the NBIA Data Retriever .)</p>
AIM Annotations (XML, zip)	 Download
Clinical Data (csv)	 Download

- A new version of RO1-023 was created to correct a cranial-caudal flip of the segmentation of the CT volume (483 images) and associated Segmentation object. The UIDs of the other scans were updated to preserve Study level consistency but were otherwise unmodified. The referenced UIDs within the AIM object for RO1-023 were updated and renamed to RO1-023v1.

- RO1-038 was updated to remove a coronal slice at the start of the of the CT volume. This created difficulty for some software to determine slice spacing.

Version 2 (Current): Updated 2017/02/28

Data Type	Download all or Query/Filter
Images (DICOM, 97.6 GB)	 (Requires the NBIA Data Retriever .)
AIM Annotations (XML, zip)	
Clinical Data (csv)	

Version 1: Updated 2015/12/22

Data Type
<p>This collection was originally submitted to TCIA as a 26 subject pilot data set. You can learn more about that subset of the collection in the following Analysis Results publication:</p> <p>NSCLC Radiogenomics: Initial Stanford Study of 26 Cases</p>