

The Clinical Proteomic Tumor Analysis Consortium Lung Adenocarcinoma Collection (CPTAC-LUAD)

Summary

Redirection Notice

This page will redirect to <https://www.cancerimagingarchive.net/collection/cptac-luad/> in about 5 seconds.

This collection contains subjects from the National Cancer Institute's [Clinical Proteomic Tumor Analysis Consortium](#) Lung Adenocarcinoma (CPTAC-LUAD) cohort. CPTAC is a national effort to accelerate the understanding of the molecular basis of cancer through the application of large-scale proteome and genome analysis, or proteogenomics. Radiology and pathology images from CPTAC patients are being collected and made publicly available by The Cancer Imaging Archive to enable researchers to investigate cancer phenotypes which may correlate to corresponding proteomic, genomic and clinical data.



Imaging from each cancer type will be contained in its own TCIA Collection, with the collection name "CPTAC-*cancertype*". Radiology imaging is collected from standard of care imaging performed on patients immediately before the pathological diagnosis, and from follow-up scans where available. For this reason the radiology image data sets are heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. Pathology imaging is collected as part of the CPTAC qualification workflow.

All CPTAC cohorts are released as either a single combined cohort, or split into *Discovery* and *Confirmatory* where applicable. There are two main types of proteomic studies: discovery proteomics and targeted proteomics. The term "discovery proteomics" is in reference to "untargeted" identification and quantification of a maximal number of proteins in a biological or clinical sample. The term "targeted proteomics" refers to quantitative measurements on a defined subset of total proteins in a biological or clinical sample, often following the completion of discovery proteomics studies to confirm interesting targets selected. Commonly used proteomic technologies and platforms are different types of mass spectrometry and protein microarrays depending on the needs, throughput and sample input requirement of an analysis, with further development on nanotechnologies and automation in the pipeline in order to improve the detection of low abundance proteins, increase throughput, and selectively reach a target protein *in vivo*. Once the protein targets of interest are identified, high-throughput targeted assays are developed for confirmatory studies: tests to affirm that the initial tests were accurate. A summary of CPTAC imaging efforts can be found on the [CPTAC Imaging Proteomics](#) page.

CPTAC Imaging Special Interest Group

You can join the [CPTAC Imaging Special Interest Group](#) to be notified of webinars & data releases, collaborate on common data wrangling tasks and seek out partners to explore research hypotheses! Artifacts from previous webinars such as slide decks and video recordings can be found on the [CPTAC SIG Webinars](#) page.

On July 22, 2020 Shankha Satpathy presented about the consortium's proteogenomic analyses of CPTAC-LUAD. This deep dive into the LUAD genomic and proteomic datasets will help researchers better understand how these can be correlated with features derived from the imaging data. ([Download the slides](#))

Acknowledgements

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

- Beaumont Health System, Royal Oak, MI - Special thanks to **George D. Wilson, PhD** from the Department of Radiation Oncology Research, **Barbara Pruetz** of the Biobank, **Debra Kapczynski, MHSA, CIIP, RT(R)(CT)** and **Rachel Deyer** from the Department of Diagnostic Radiology.
- Boston Medical Center, Boston, MA - Special thanks to **Chris D. Andry M.Phil, PhD** from the Department of Pathology and Laboratory Medicine, **Margaret Lavoye, Artem Kaliaev, Wilson Chavez, Stephan Anderson, Jorge Soto**, and **Mitchell Horn** from the Department of Radiology, **Elizabeth Duffy, MA** and **Cheryl Spencer, MA** of the Biobank.
- St. Joseph's Hospital and Medical Center, Phoenix, AZ - Special thanks to **Jennifer Eschbacher, MD** from the Department of Neuropathology, **Catherine Seiler, PhD**, **Rosy Singh** and **Beth Hermes** from the Biobank Core Facility, and **Victor Sisneros, RT(R)(CT), CPSA**.
- International Institute for Molecular Oncology, Pozna, Poland - Special thanks to **Maciej Wiznerowicz MD, PhD** and **Jan Lubiski MD PhD**; **Wojciech Dyszkiewicz MD, PhD.**, and **ukasz Gsiorowski MD** from Greater Poland Center of Pulmonology and Thoracic Surgery, Pozna, Poland; **Bartosz Kubisa MD, PhD.**, **Malgorzata Wojtys MD and PhD.**, and **Tomasz Grodzki MD, PhD** of Regional Hospital Complex, Szczecin, Poland
- BioPartners, CA - Special thanks to **Alexander Gasparian, PhD.** from the Department of Drug Discovery and Biomedical Sciences, University of South Carolina College of Pharmacy, **Kakhaber Zaalishvili, MD** Medical Advisor and Staff Pathologist at BioPartners, LLC, **Milla Gorodnia**, President of BioPartners, Inc., **Victoria Christensen**, Global Business Development/Project Coordination Manager, **Oksana Havryliuk, MD.** Chief of Research department of radiodiagnostics of NCI (Ukraine), **Marianna Gredil'**, Director of BioPartners, LLC, and **Anna Legenka** Chief of the Data Department at BioPartners, LLC

Data Access

Data Access

Data Type	Download all or Query/Filter	License
Images (DICOM, 28.2 GB)	Download Search (Download requires the NBIA Data Retriever)	CC BY 3.0
Tissue Slide Images (SVS, 425 GB)	Download Search (Download and apply the IBM-Aspera-Connect plugin to your browser to retrieve this faspex package)	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)
- [Proteomic Data Commons \(PDC\)](#) (Proteomic & Clinical Data)
- [Genomic Data Commons \(GDC\)](#) (Genomic & Clinical Data)

Detailed Description

Detailed Description

	Radiology Image Statistics	Pathology Image Statistics
Modalities	CT, MR, PT	Pathology
Number of Participants	25	244
Number of Studies	38	N/A
Number of Series	242	N/A
Number of Images	43,333	1,139
Images Size (GB)	28.2	425

Accessing CPTAC publication cohorts

All CPTAC cohorts are released as either a single combined cohort, or split into *Discovery* and *Confirmatory* where applicable. In the case of CPTAC-LUAD there was a "Discovery Cohort" release. Images associated with these cases can be downloaded using the following links:

- [CPTAC-LUAD Radiology Discovery Cohort](#)
- [CPTAC-LUAD Pathology Discovery Cohort](#)

Accessing the Proteomic & Genomic Clinical Data

To access/download the clinical data on the **Proteomic Data Commons (PDC)** and **Genomic Data Commons (GDC)**, once you have identified the data of your interest, move to the 'Clinical' tab on the browse page. Select the checkbox to select a specific row, all rows on the page or all pages and click the export clinical manifest button in CSV or TSV format on the GDC, or TSV or JSON format on the PDC.

A Note about TCIA and CPTAC Subject Identifiers and Dates

Subject Identifiers:

A subject with radiology and pathology images stored in TCIA is identified with a de-identified project Patient ID that is identical to the Patient ID of the same subject with clinical, proteomic, and/or genomic data stored in other CPTAC databases and web sites.

Dates:

The radiology imaging data is in DICOM format. To provide temporal context information aligned with events in the clinical data set for each patient, TCIA has inserted information in DICOM tag (0012,0050) *Clinical Trial Time Point ID*. This DICOM tag contains the number of days from the date the patient was initially diagnosed pathologically with the disease to the date of the scan. E.g. a scan acquired 3 days before the diagnosis would contain the value -3. A follow up scan acquired 90 days after diagnosis would contain the value 90.

The DICOM date tags (i.e. birth dates, imaging study dates, etc.) are modified per TCIA's standard process which offsets them 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.

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

National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC). (2018). **The Clinical Proteomic Tumor Analysis Consortium Lung Adenocarcinoma Collection (CPTAC-LUAD)** (Version 12) [Data set]. The Cancer Imaging Archive. <https://doi.org/10.7937/K9/TCIA.2018.PAT12TBS>

Acknowledgement

The CPTAC program requests that publications using data from this program include the following statement: **"Data used in this publication were generated by the National Cancer Institute Clinical Proteomic Tumor Analysis Consortium (CPTAC)."**

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 TCIA data. If you have a manuscript you'd like to add please [contact TCIA's Helpdesk](#).

Versions

Version 12 (Current): Updated 2023/02/24

Data Type	Download all or Query/Filter
Images (DICOM, 28.2 GB)	Download Search (Requires NBIA Data Retriever .)
Tissue Slide Images (SVS, 425 GB)	Download Search

Proteomics (web)	Search
------------------	------------------------

Radiology modality data cleanup to remove extraneous scans.

Version 11: Updated 2020/09/03

Data Type	Download all or Query/Filter
Images (DICOM, 28.9 GB)	Download Search (Requires NBIA Data Retriever.)
Tissue Slide Images (SVS, 425 GB)	Download Search
Proteomics (web)	Search

Changed to new Aspera download link for histopathology slides.

Version 10: Updated 2020/03/31

Data Type	Download all or Query/Filter
Images (DICOM, 28.9 GB)	Download Search (Requires NBIA Data Retriever.)
Tissue Slide Images (SVS, 425 GB)	Search
Proteomics (web)	Search

Added 1 radiology subject

Version 9 : Updated 2020/02/14

Data Type	Download all or Query/Filter
Images (DICOM, 25 GB)	Download (Requires NBIA Data Retriever.)
Tissue Slide Images (SVS, 425 GB)	Search
Proteomics (web)	Search

Added 14 new pathology subjects and 38 new pathology images

Version 8: Updated 2019/12/19

Data Type	Download all or Query/Filter
Images (DICOM, 25 GB)	Download (Requires NBIA Data Retriever.)

Tissue Slide Images (SVS, 390.4 GB)	Search
Proteomics (web)	Search

Added 5 new radiology subjects.

Version 7: Updated 2019/09/30

Data Type	Download all or Query/Filter
Images (DICOM, 17.9 GB)	Download (Requires NBIA Data Retriever .)
Tissue Slide Images (SVS, 390.4 GB)	
Proteomics (web)	Search

Added new subjects.

Version 6: Updated 2019/06/30

Data Type	Download all or Query/Filter
Images (DICOM, 12.2 GB)	Download (Requires NBIA Data Retriever .)
Tissue Slide Images (SVS, 390.4 GB)	
Proteomics (web)	Search

Added Subjects

Version 5: Updated 2019/03/31

Data Type	Download all or Query/Filter
Images (DICOM, 11.4 GB)	Download (Requires NBIA Data Retriever .)
Tissue Slide Images (web)	Search
Proteomics (web)	Search

Added new subjects

Version 4 : Updated 2018/10/24

Data Type	Download all or Query/Filter
Images (DICOM, 3.8 GB)	Download (Requires NBIA Data Retriever .)
Tissue Slide Images (web)	Search
Proteomics (web)	Search

Added subjects

Version 3: Updated 2018/06/30

Data Type	Download all or Query/Filter
Images (DICOM, 4.0 GB)	Download (Requires NBIA Data Retriever .)
Tissue Slide Images (web)	Search
Proteomics (web)	Search

Added new subjects.

Version 2: Updated 2018/04/26

Data Type	Download all or Query/Filter
Images (DICOM, 3.9 GB)	Download (Requires NBIA Data Retriever .)
Tissue Slide Images (web)	Search
Proteomics (web)	Search

Added new subjects.

Version 1: Updated 2018/01/10

Data Type	Download all or Query/Filter
Images (DICOM, 2.77 GB)	Download (Requires NBIA Data Retriever .)

Tissue Slide Images (web)	Search
Proteomics (web)	Search