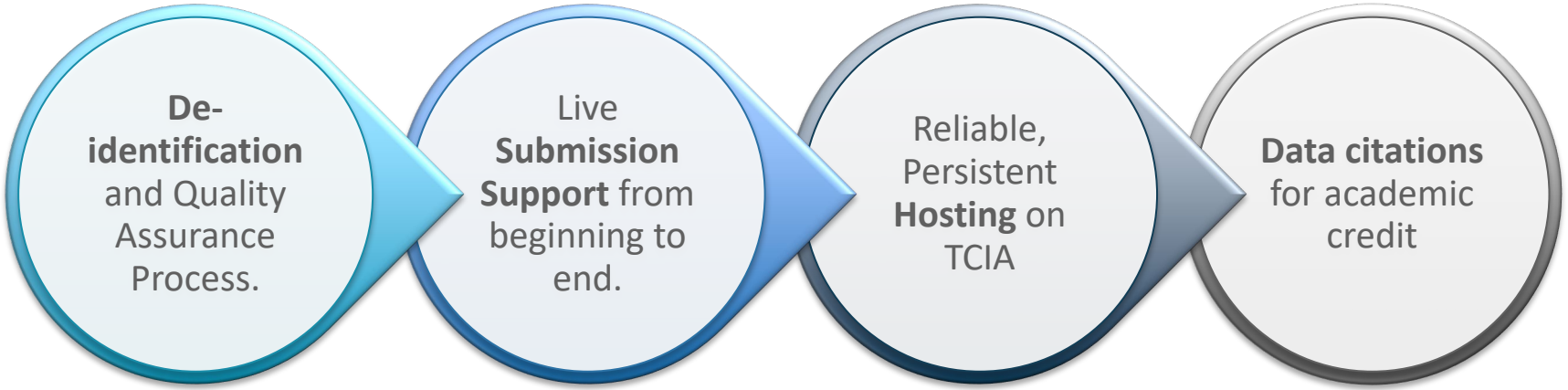
Support for the Clinical Proteomic Tumor Analysis Consortium

Justin Kirby (justin.kirby@nih.gov)
Technical Director, Cancer Imaging Informatics Lab
Frederick National Laboratory for Cancer Research

NIH NATIONAL CANCER INSTITUTE

TCIA's Mission

Provide services to encourage data sharing for cancer imaging research



TCIA components

Data Collection Centers

- Tools and staffing to support data collection, curation, and de-identification

Data Access Tools*

- Browse
- Filter/Search
- Visualize
- REST API

Data Analysis Centers

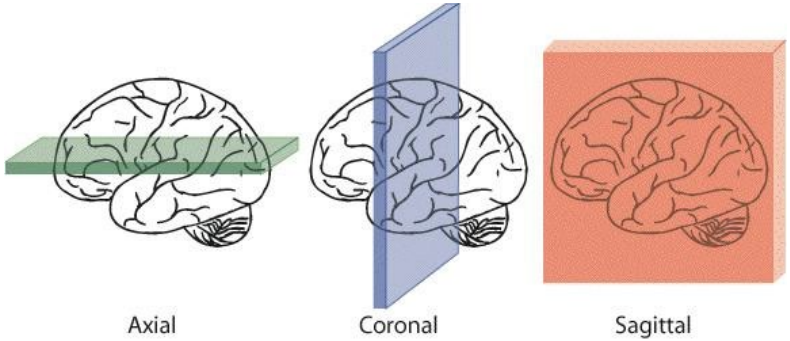
- 3rd party web sites or tools which use TCIA's API or mirror its data

* Significant support from NCI CBIIT and ITCR programs

Radiology Imaging Overview

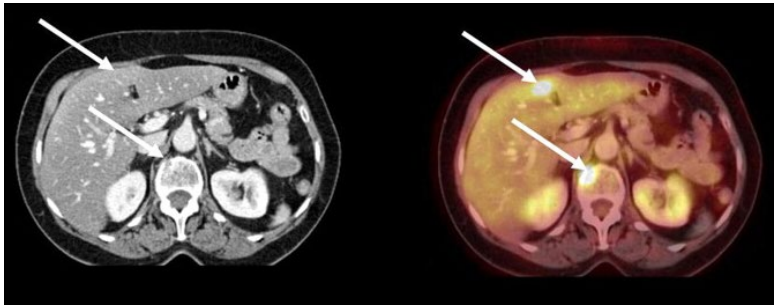
- When a patient comes in for imaging studies it's common to use:

Different angles



Top to bottom | Front to Back | Side to Side

Different modalities

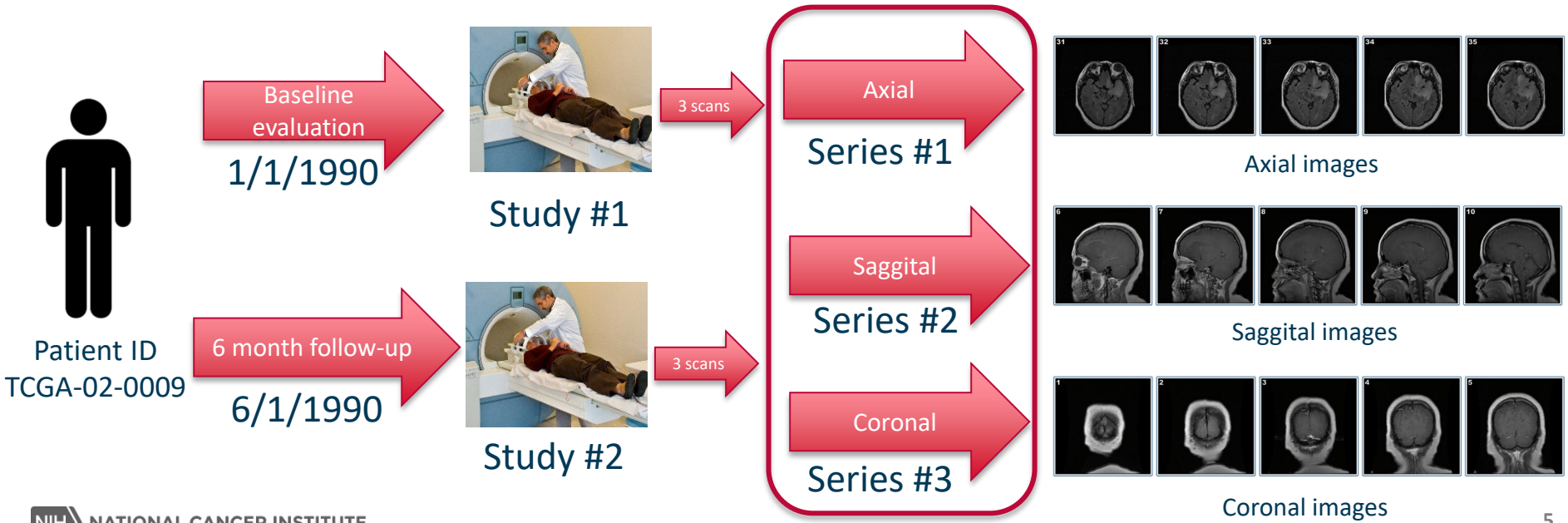


Computed Tomography (CT)

Positron emission tomography (PET)

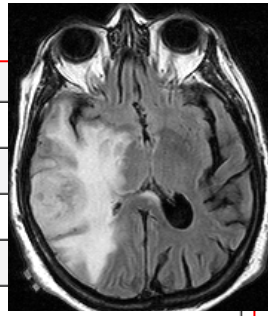
Radiology Imaging Format: DICOM

- The DICOM data is hierarchical in its structure
- Patient → Study (appointment) → Series (scans) → Images



Tackling the de-identification challenge

- PHI can appear in hundreds of places in DICOM
  - Dates
  - Identifiers
  - Descriptions



(0008,0022)	Acquisition Date	19970614
(0008,0023)	Content Date	19970614
	Study Time	073837
(0008,0031)	Series Time	074251
(0008,0032)	Acquisition Time	074252
(0008,0033)	Content Time	074252
(0008,0050)	Accession Number	2819497684894126
	Modality	MR
(0008,0070)	Manufacturer	GE MEDICAL SYSTEMS
(0008,0090)	Referring Physician's Name	
(0008,1010)	Station Name	
(0008,1030)	Study Description	MRI, BRAIN W&W/O CONTRAMR
	Procedure Code Sequence	
(0008,103E)	Series Description	AX FLAIR
(0008,1090)	Manufacturer's Model Name	GENESIS_SIGNA
(0010,0010)	Patient's Name	
(0010,0020)	Patient ID	TCGA-02-0009
	Patient's Birth	

Tackling the de-identification challenge

Documents saved as images

(e.g. billing, patient records)

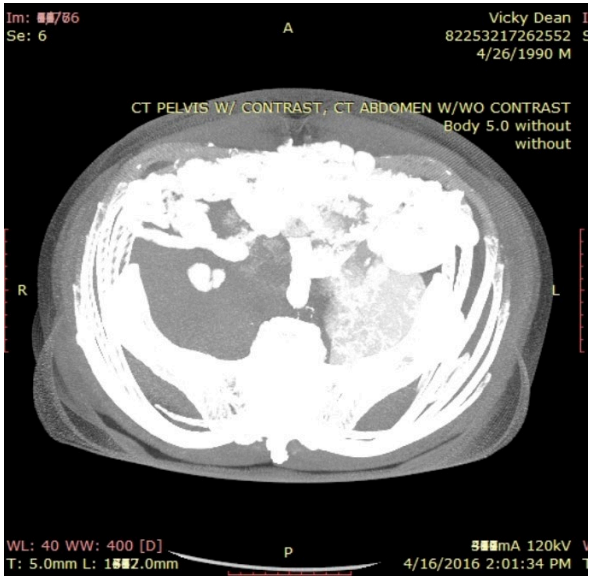
November 2010

Paper Into PACS
By Dan Harvey
Radiology Today
Vol. 11 No. 11 P. 20

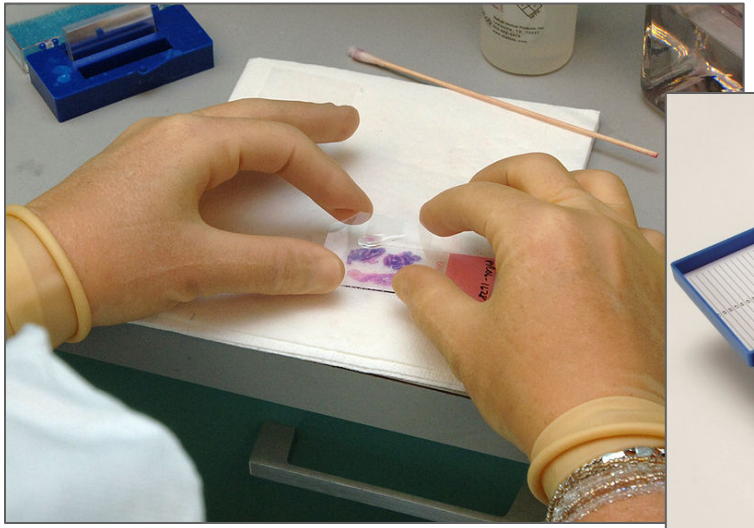
Converting paper documents into a format acceptable to store in a PACS is a widespread need. Imaging facilities use several approaches to solve this common problem, including using DICOM, RIS, or EMR document management.

PHI in the image pixels

(fake example)



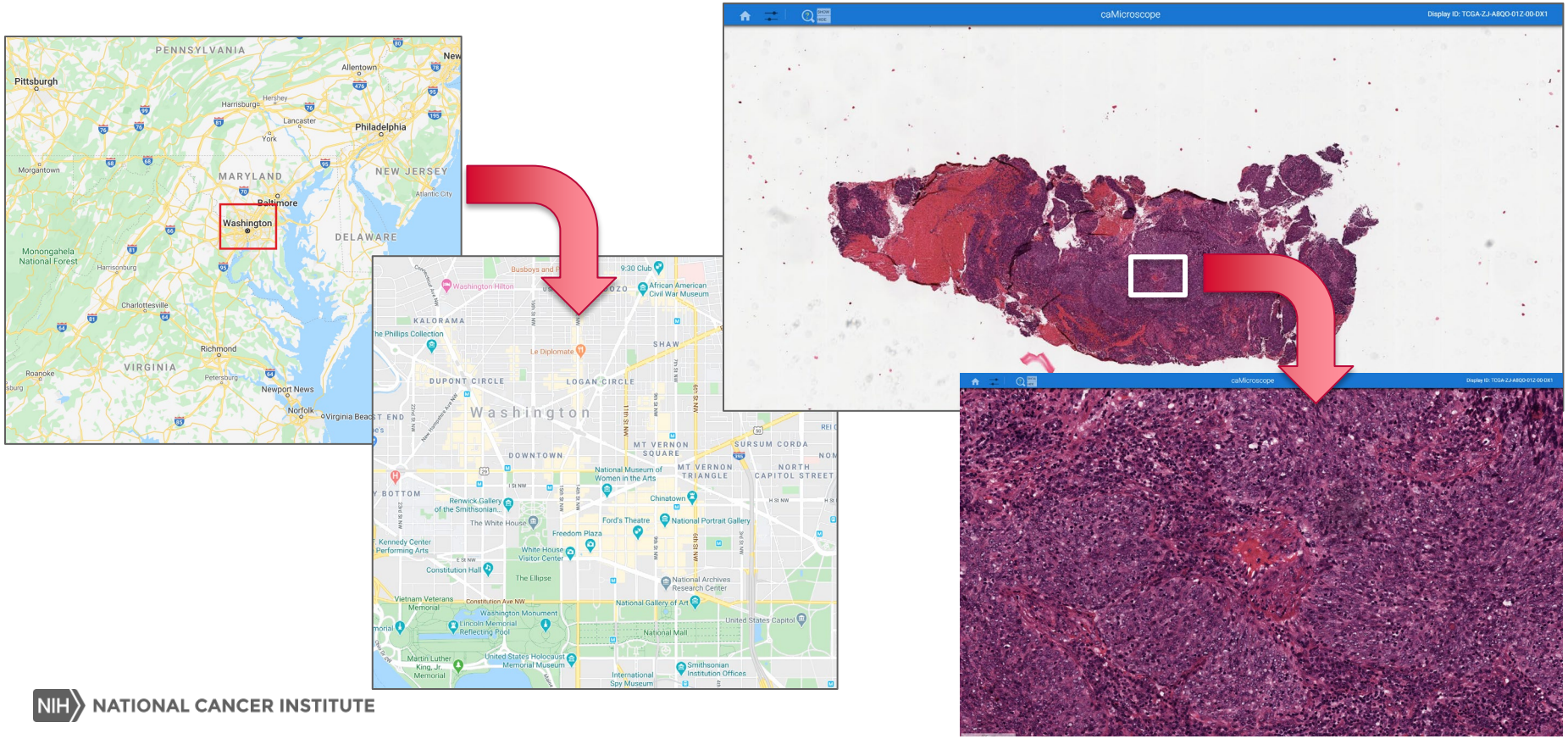
Histopathology Imaging Overview



Stained tissue mounted on glass slide

Slides being digitized

Histopathology Imaging Overview



Histopathology image de-identification

- Image metadata
- Image labels/barcodes
- Pixel data

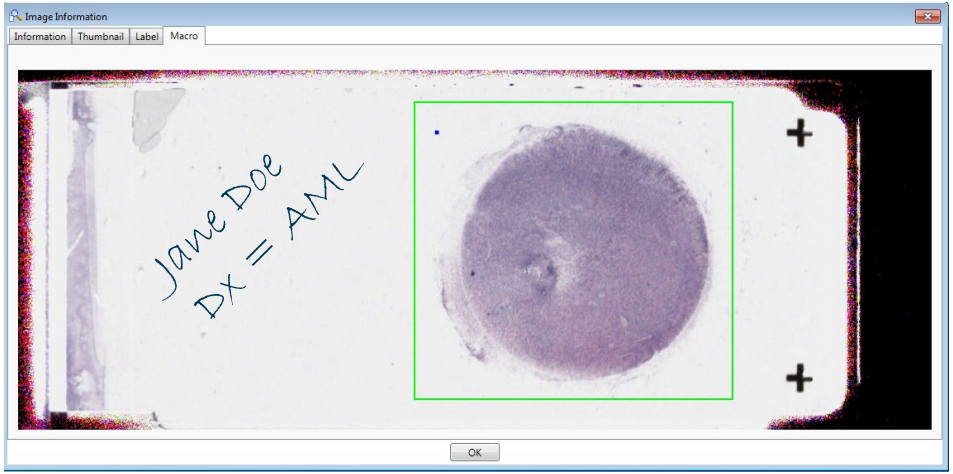


Image Information

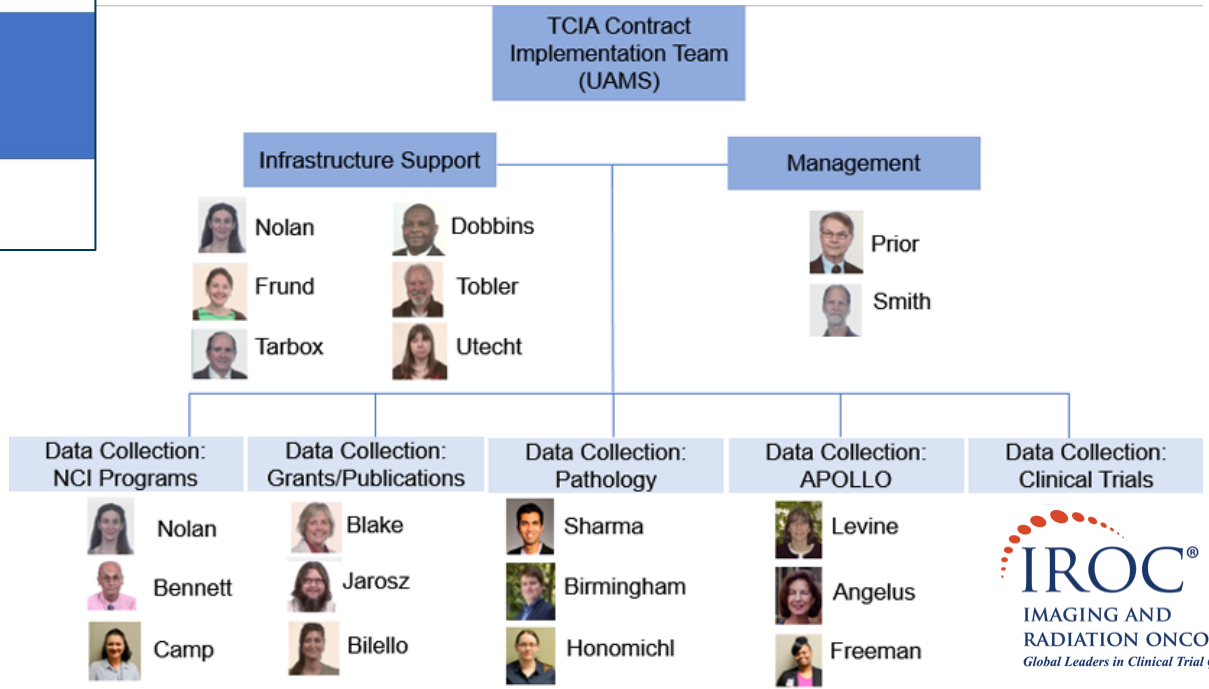
Information | Thumbnail | Label | Macro

Image Path	\\vmware-host\Shared Folders\Downloads\C3L-00452-41.svs
Description	Aperio Image Library v12.0.11 96224x85778 [0,100 92344x85678] (240x240) JPEG/RGB Q=70
Apparent Magnification	40X
StripeWidth	992
ScanScope ID	SS1553
Filename	104632
Date	01/25/17
Time	12:32:38
Time Zone	GMT-05:00
User	bd156663-fd1a-483b-a2be-9aa5f37aae07
MPP	0.2467
Left	28.078907
Top	23.057886
LineCameraSkew	0.001659
LineAreaXOffset	0.011789
LineAreaYOffset	-0.003588
Focus Offset	0.000000
DSR ID	aperio04
ImageID	104632
Exposure Time	109
Exposure Scale	0.000001
DisplayColor	0
OriginalWidth	96224
OriginalHeight	85778
ICC Profile	ScanScope v1
Image Type	SVS/JPEG 2
Image Width	92,344 pixels
Image Height	85,678 pixels
Image Depth	1 pixels
Image Channels	3
Image Bit Depth	8 bits
File Size	2,257,398,000 bytes
Compression Type	JPEG using libjpeg
Compression Quality	70
Compression Ratio	10.51
Organization	tiled
Tile Width	240 pixels

Pyramid	Width	Height	Ratio
Thumbnail	827	768	111:1
Level 3	2885	2677	32:1
Level 2	5771	5354	16:1
Level 1	23086	21419	4:1
Base	92344	85678	

TCIA Staffing: De-id and curation is a major effort

- 5 curation teams
- Infrastructure support
- Management

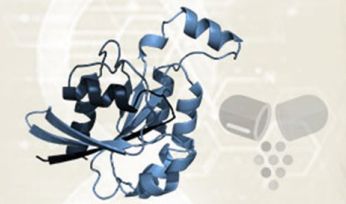


IROC®
IMAGING AND RADIATION ONCOLOGY CORE
Global Leaders in Clinical Trial Quality Assurance

TCIA Data Sources

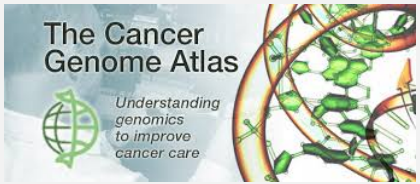
NCI data collection initiatives

CLINICAL PROTEOMIC TUMOR ANALYSIS CONSORTIUM



The Cancer Genome Atlas

Understanding genomics to improve cancer care

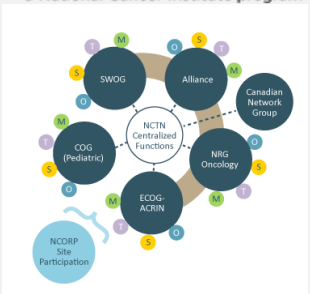


APOLLO

Applied Proteogenomics OrganizationAL Learning And Outcomes Consortium

NCI National Clinical Trials Network

a National Cancer Institute program

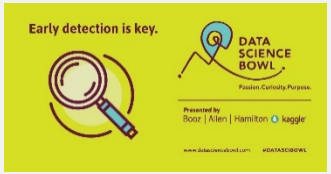


"Community" proposals

reviewed monthly

NIH All About Grants

Data generated by NCI/NIH Grants



Challenge competitions

SCIENTIFIC DATA

PLOS ONE

Publication data sharing requests

Data Sources: Analyses of TCIA Collections

nature > scientific data > data descriptors > article

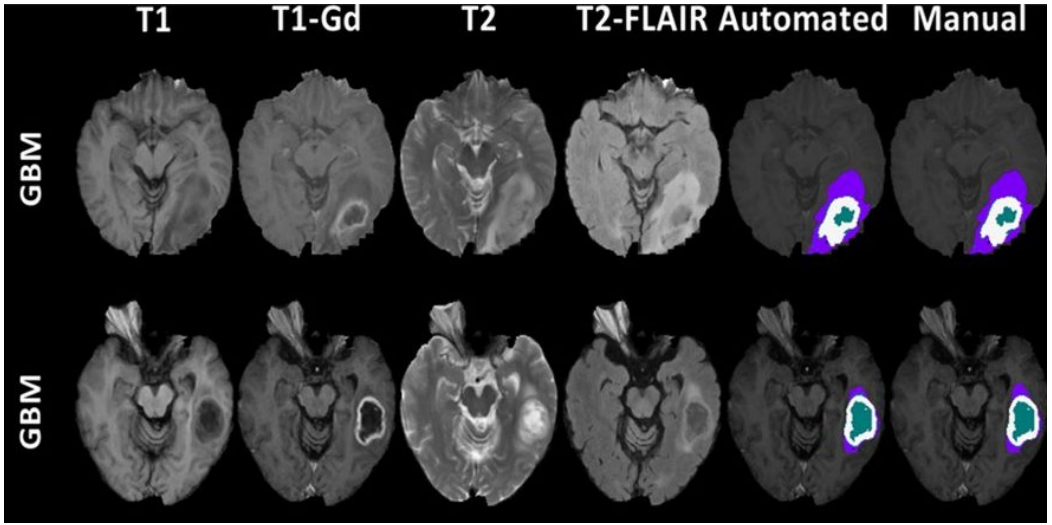
SCIENTIFIC DATA

Data Descriptor | OPEN | Published: 05 September 2017

Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features

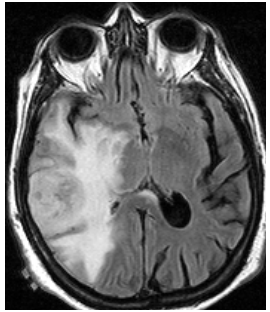
Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S. Kirby, John B. Freymann, Keyvan Farahani & Christos Davatzikos

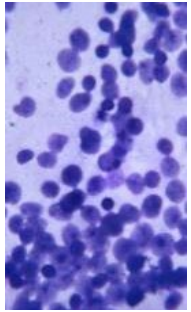
Scientific Data **4**, Article number: 170117 (2017) | Download Citation

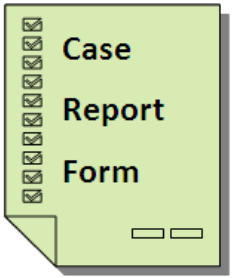


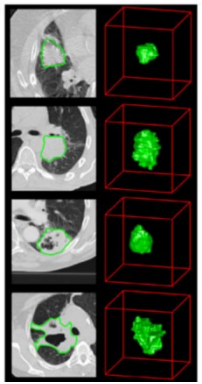
Example analysis data derived from existing TCGA-GBM collection

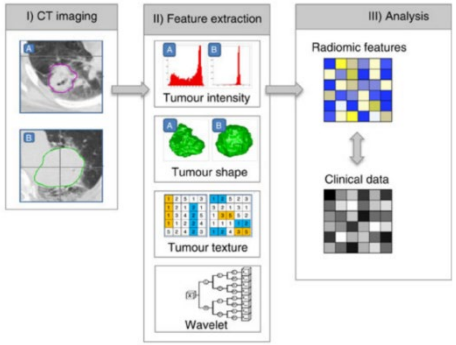
TCIA Data Summary









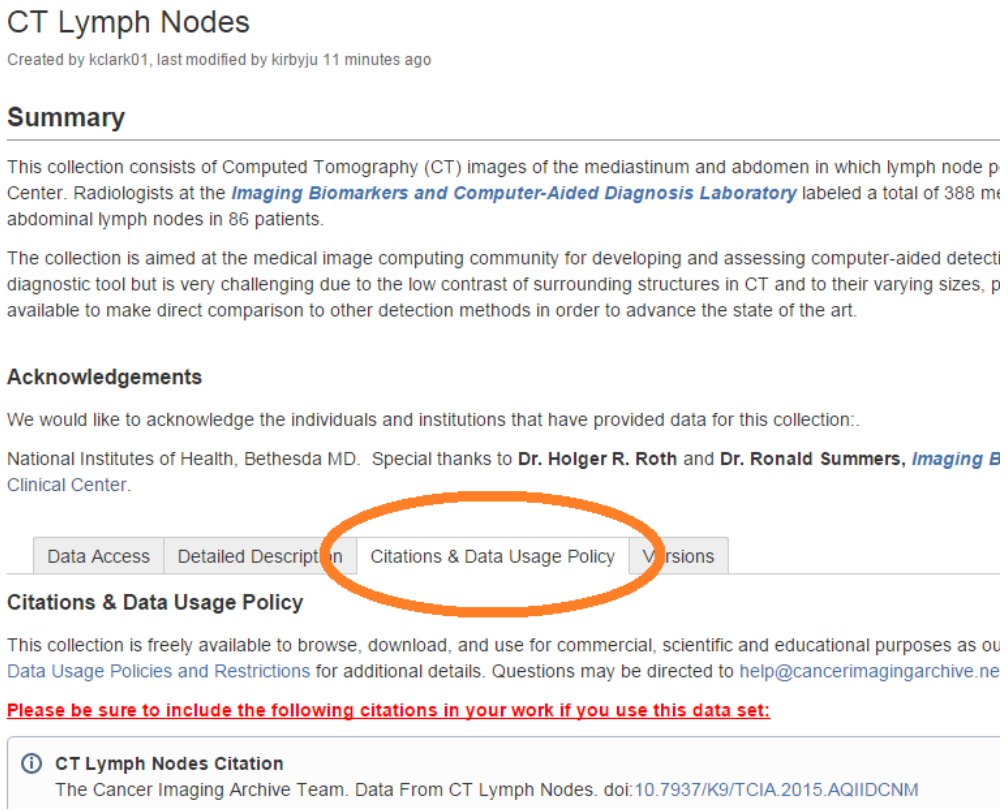


http://www.cancerimagingarchive.net

- New collection proposals are reviewed by the TCIA Advisory Group for quality/utility
- 115 collections consisting of ~47,300 subjects available for download
- Covers radiology, radiation therapy, and pathology image modalities
- Wide variety of cancers + phantoms
- Most have associated supporting data
  - Demographics/outcomes/therapy
  - Image analyses (annotations, segmentations, features)
  - Genomics/Proteomics

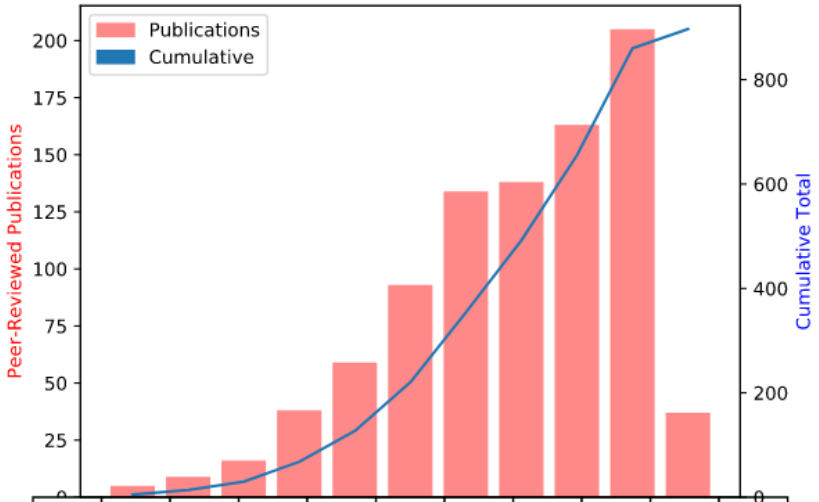
Citations and data usage policy

- Citations (credit for data sharing!)
  - Dataset
  - Scientific manuscript
  - TCIA manuscript
- Data Usage Policy
  - Generally 100% open (no login)
  - 11 datasets require special access requests
  - Creative Commons Attribution license



TCIA Metrics

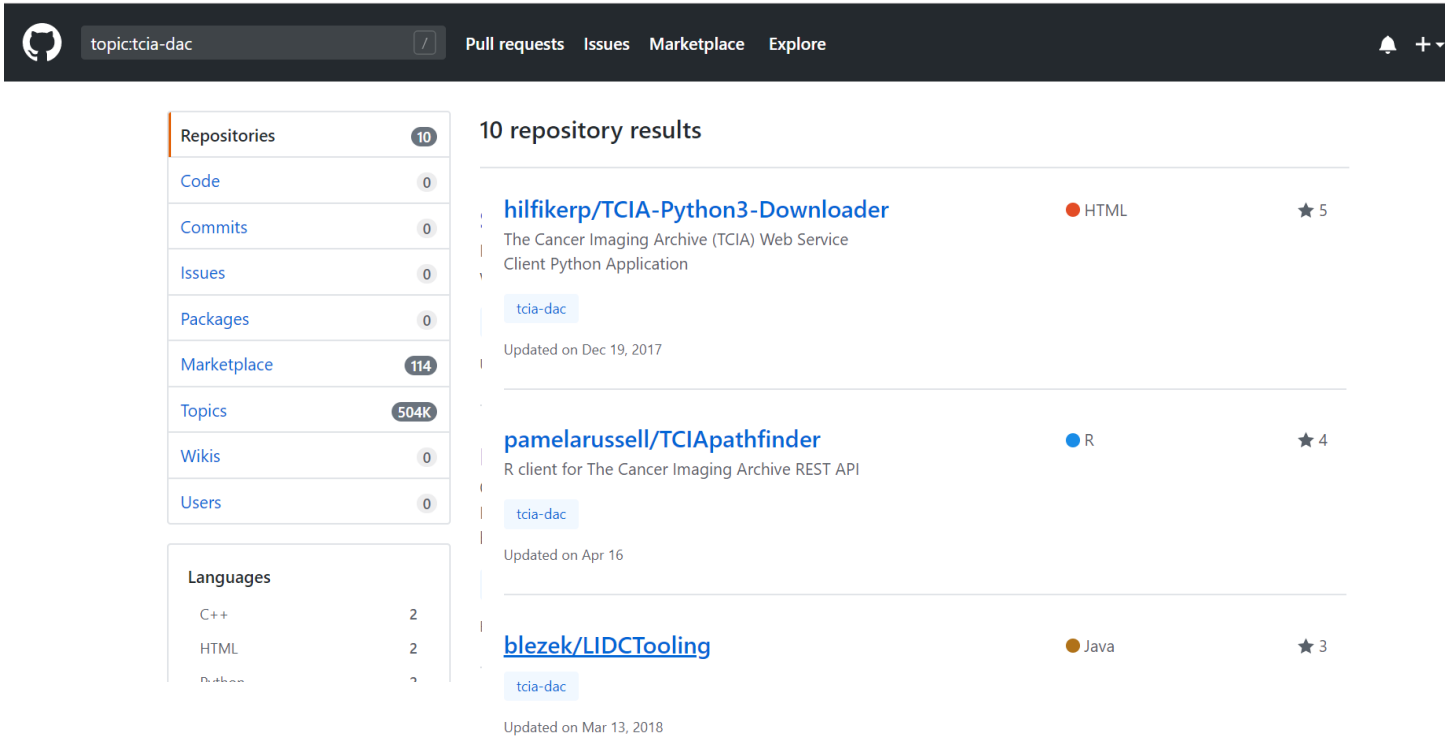
- Over 20,000 active users per month
- Downloads of ~100TB per month
- 30 incoming data sets to be curated
- 897 peer reviewed publications based on TCIA
- Widely used by industry due to permissive Creative Commons licensing



	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Publications	5	9	16	38	59	93	134	138	163	205	37
Cumulative	5	14	30	68	127	220	354	492	655	860	897

TCIA community code sharing

Tag your Github repo with “tcia-dac” topic tag to appear in the list

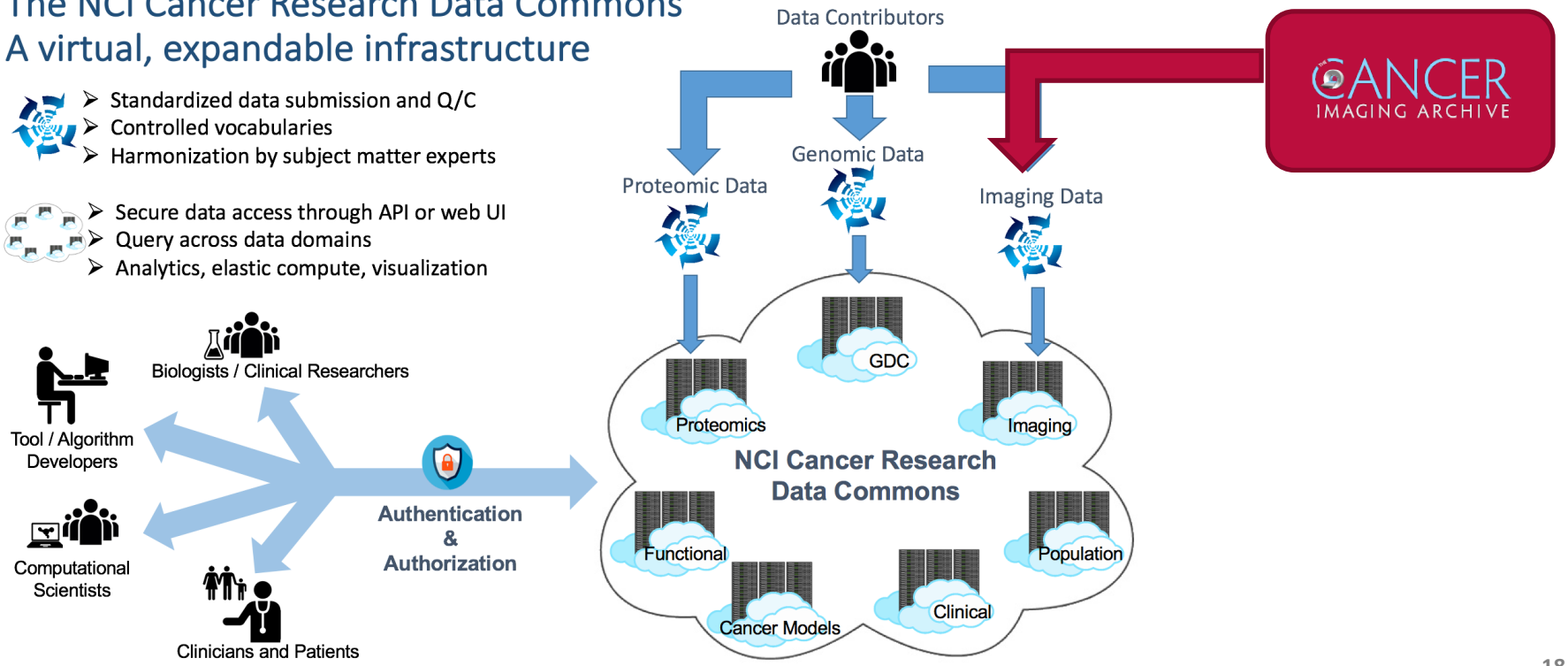


TCIA data in the NCI Data Commons cloud

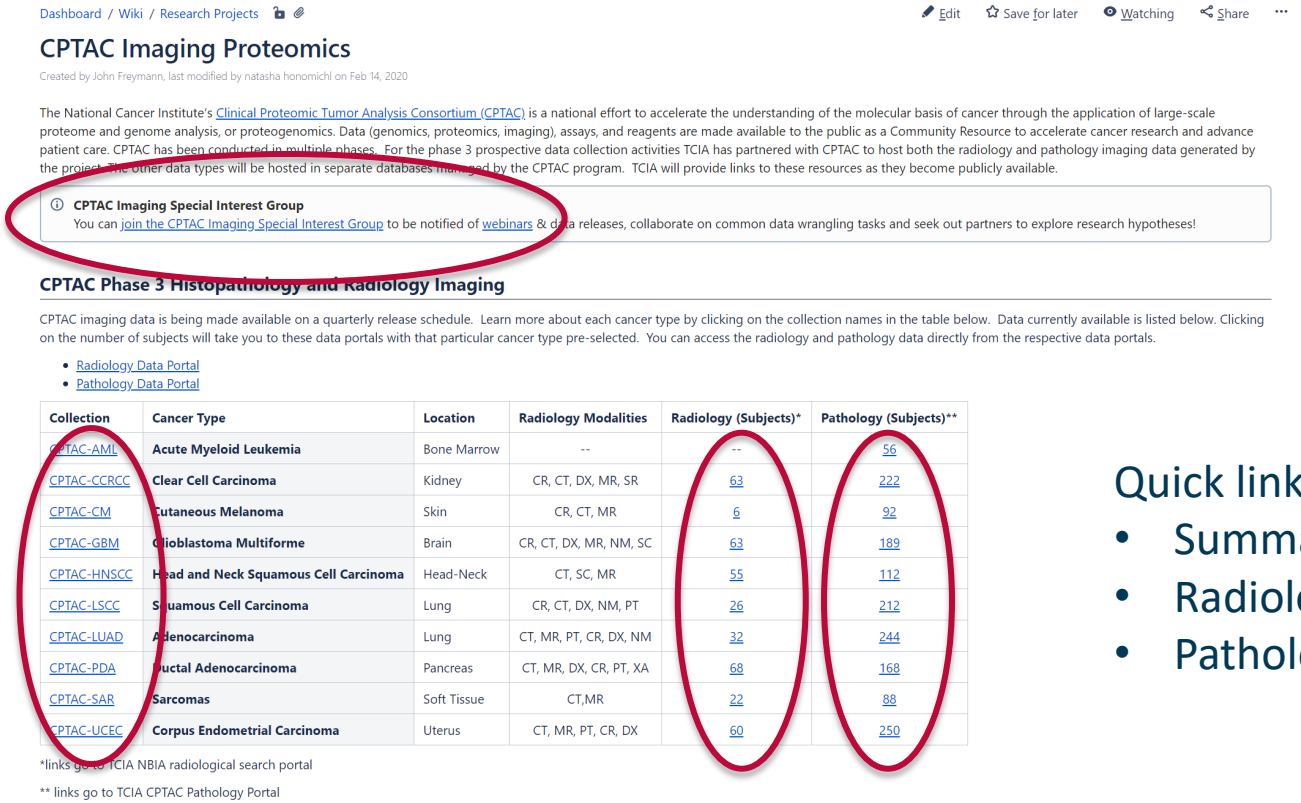
The NCI Cancer Research Data Commons
A virtual, expandable infrastructure

- Standardized data submission and Q/C
- Controlled vocabularies
- Harmonization by subject matter experts

- Secure data access through API or web UI
- Query across data domains
- Analytics, elastic compute, visualization



TCIA Support for CPTAC: Project Summary



Dashboard / Wiki / Research Projects

Edit Save for later Watching Share

CPTAC Imaging Proteomics

Created by John Freymann, last modified by natasha honomichl on Feb 14, 2020

The National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (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. Data (genomics, proteomics, imaging), assays, and reagents are made available to the public as a Community Resource to accelerate cancer research and advance patient care. CPTAC has been conducted in multiple phases. For the phase 3 prospective data collection activities TCIA has partnered with CPTAC to host both the radiology and pathology imaging data generated by the project. The other data types will be hosted in separate databases maintained by the CPTAC program. TCIA will provide links to these resources as they become publicly available.

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!

CPTAC Phase 3 Histopathology and Radiology Imaging

CPTAC imaging data is being made available on a quarterly release schedule. Learn more about each cancer type by clicking on the collection names in the table below. Data currently available is listed below. Clicking on the number of subjects will take you to these data portals with that particular cancer type pre-selected. You can access the radiology and pathology data directly from the respective data portals.

- Radiology Data Portal
- Pathology Data Portal

Collection	Cancer Type	Location	Radiology Modalities	Radiology (Subjects)*	Pathology (Subjects)**
CPTAC-AML	Acute Myeloid Leukemia	Bone Marrow	--	--	56
CPTAC-CCRCC	Clear Cell Carcinoma	Kidney	CR, CT, DX, MR, SR	63	222
CPTAC-CM	Cutaneous Melanoma	Skin	CR, CT, MR	6	92
CPTAC-GBM	Glioblastoma Multiforme	Brain	CR, CT, DX, MR, NM, SC	63	189
CPTAC-HNSCC	Head and Neck Squamous Cell Carcinoma	Head-Neck	CT, SC, MR	55	112
CPTAC-LSCC	Squamous Cell Carcinoma	Lung	CR, CT, DX, NM, PT	26	212
CPTAC-LUAD	Adenocarcinoma	Lung	CT, MR, PT, CR, DX, NM	32	244
CPTAC-PDA	Ductal Adenocarcinoma	Pancreas	CT, MR, DX, CR, PT, XA	68	168
CPTAC-SAR	Sarcomas	Soft Tissue	CT,MR	22	88
CPTAC-UCEC	Corpus Endometrial Carcinoma	Uterus	CT, MR, PT, CR, DX	60	250

*links go to TCIA NBIA radiological search portal

** links go to TCIA CPTAC Pathology Portal

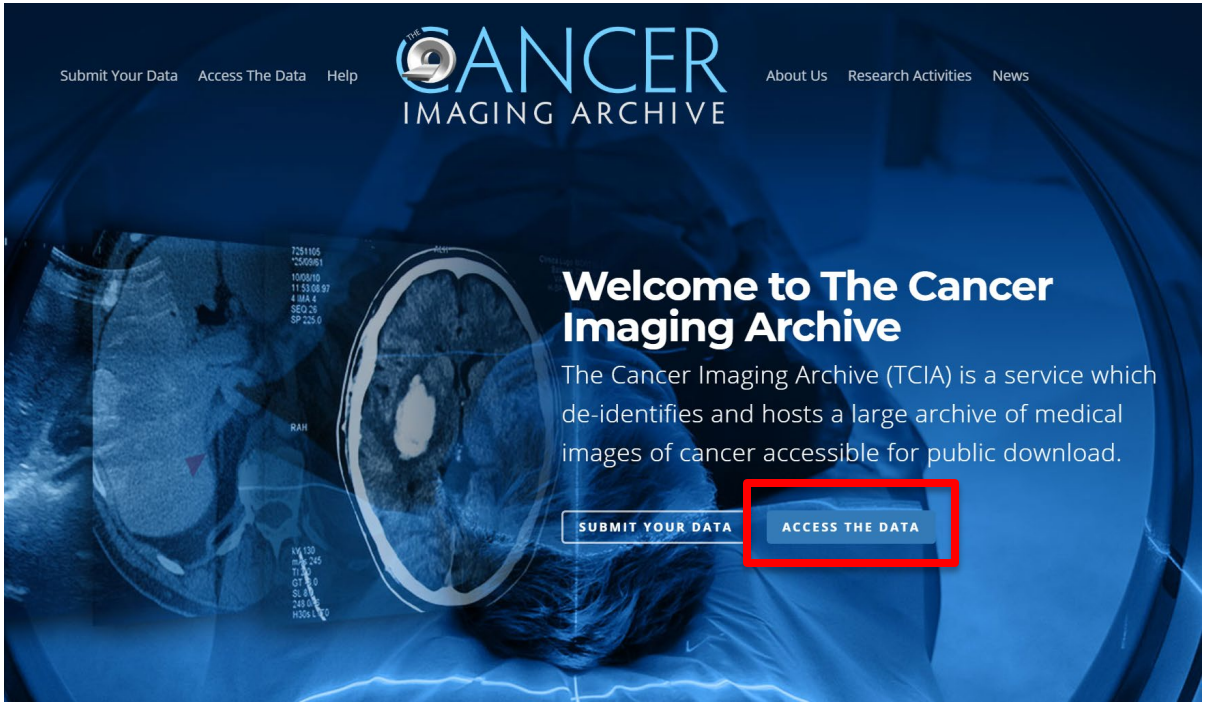
Quick introduction to the TCIA CPTAC Pathology Portal:

Quick links:

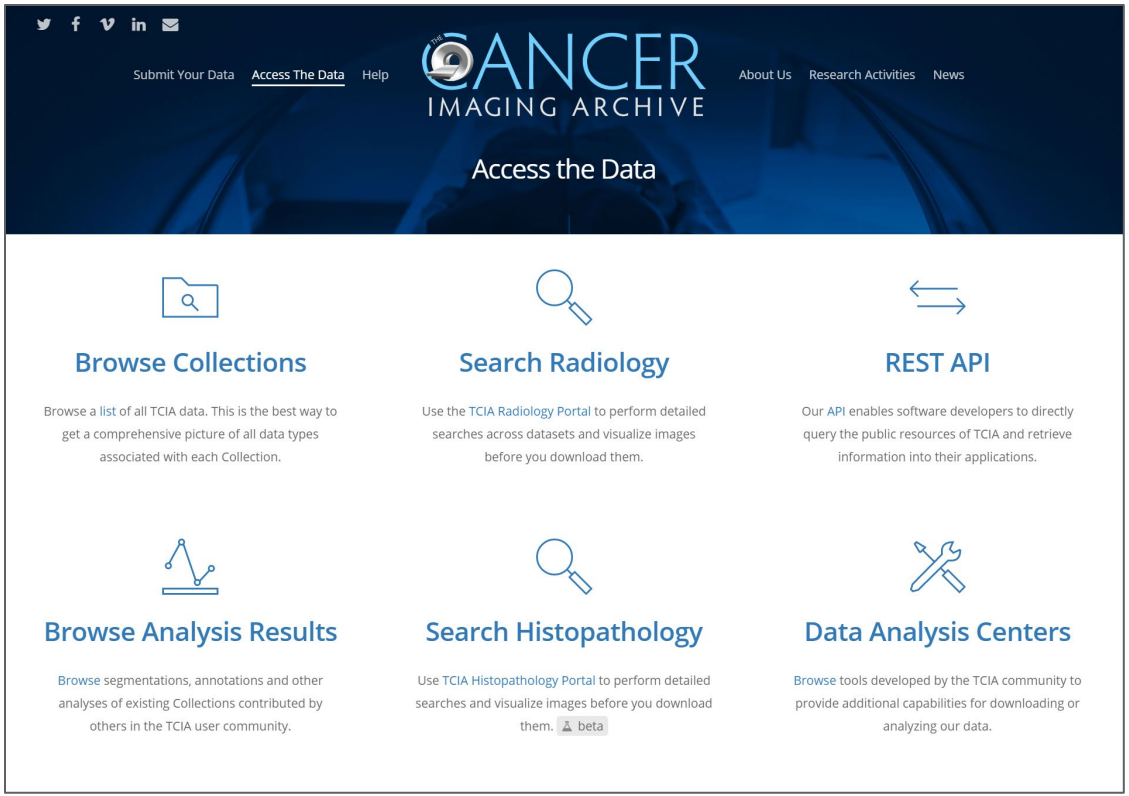
- Summary pages
- Radiology Portal
- Pathology Portal

https://wiki.cancerimagingarchive.net/x/mIREAQ

Accessing the Data



Accessing the Data



Radiology Downloads & Data Portal

CPTAC-GBM

Created by Tracy Nolan, last modified by natasha honomichl on Feb 14, 2020

Summary

This collection contains subjects from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme (CPTAC-GBI molecular basis of cancer through the application of large-scale proteome and genome analysis, or proteogenomics. Radiology and pathology images fr by The Cancer Imaging Archive to enable researchers to investigate cancer phenotypes which may correlate to corresponding proteomic, genomic and cl

CPTAC Phase 3 collects data from ten cancer types. In TCIA, imaging from each cancer type will be contained in its own TCIA Collection, with the collectic available on TCIA each quarter as it is collected. A summary of CPTAC Phase 3 imaging efforts can be found on the CPTAC Imaging Proteomics page.

Radiology imaging is collected from standard of care imaging performed on patients immediately before the pathological diagnosis, and from follow-up heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. Pathology imaging is collected as part of the CPTAC qualificatior

CPTAC Imaging Special Interest Group

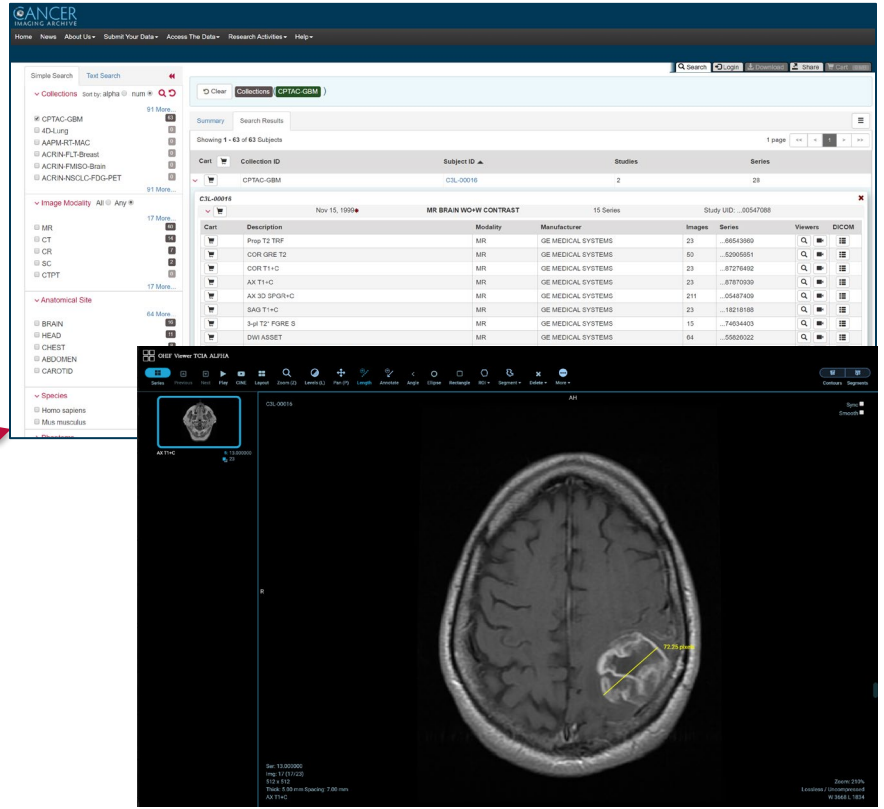
You can join the CPTAC Imaging Special Interest Group to be notified of webinars & data releases, collaborate on common data wrangling tasks and seek webinars such as slide decks and video recordings can be found on the CPTAC SIG Webinars page.

Data Access | Detailed Description | Citations & Data Usage Policy | Versions

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 t download a subset of its contents.

Data Type	Download all or Query/Filter
Images (DICOM, 39.4 GB)	Download Search
Tissue Slide Images (SVS, 87 GB)	Download Search
Clinical Data API (JSON - more info)	Download
Discovery Study Proteomics/Clinical Data (external)	• CPTAC Data Portal (Georgetown) • Proteomic Data Commons
Genomics/Clinical Data (External)	Genomic Data Commons



Pathology Downloads & Data Portal

CPTAC-GBM

Created by Tracy Nolan, last modified by natasha honomichl on Feb 14, 2020

Summary

This collection contains subjects from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme (CPTAC-GBI molecular basis of cancer through the application of large-scale proteome and genome analysis, or proteogenomics. Radiology and pathology images fr by The Cancer Imaging Archive to enable researchers to investigate cancer phenotypes which may correlate to corresponding proteomic, genomic and cl

CPTAC Phase 3 collects data from ten cancer types. In TCIA, imaging from each cancer type will be contained in its own TCIA Collection, with the collectic available on TCIA each quarter as it is collected. A summary of CPTAC Phase 3 imaging efforts can be found on the CPTAC Imaging Proteomics page.

Radiology imaging is collected from standard of care imaging performed on patients immediately before the pathological diagnosis, and from follow-up heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. Pathology imaging is collected as part of the CPTAC qualificatior

CPTAC Imaging Special Interest Group

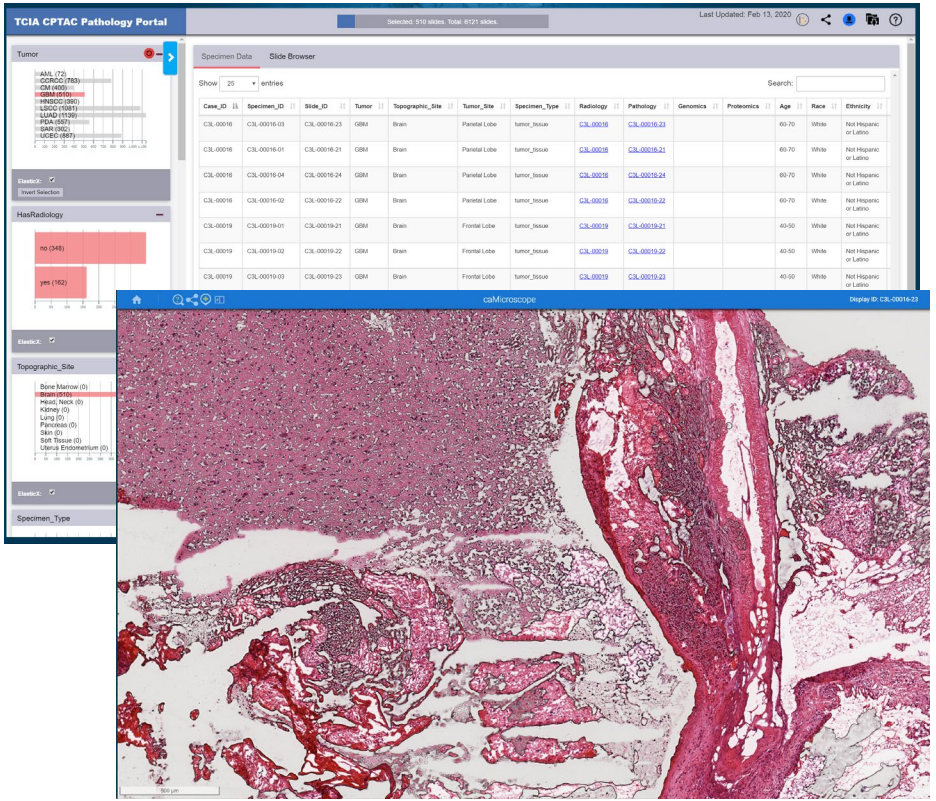
You can join the CPTAC Imaging Special Interest Group to be notified of webinars & data releases, collaborate on common data wrangling tasks and seek webinars such as slide decks and video recordings can be found on the CPTAC SIG Webinars page.

Data Access | Detailed Description | Citations & Data Usage Policy | Versions

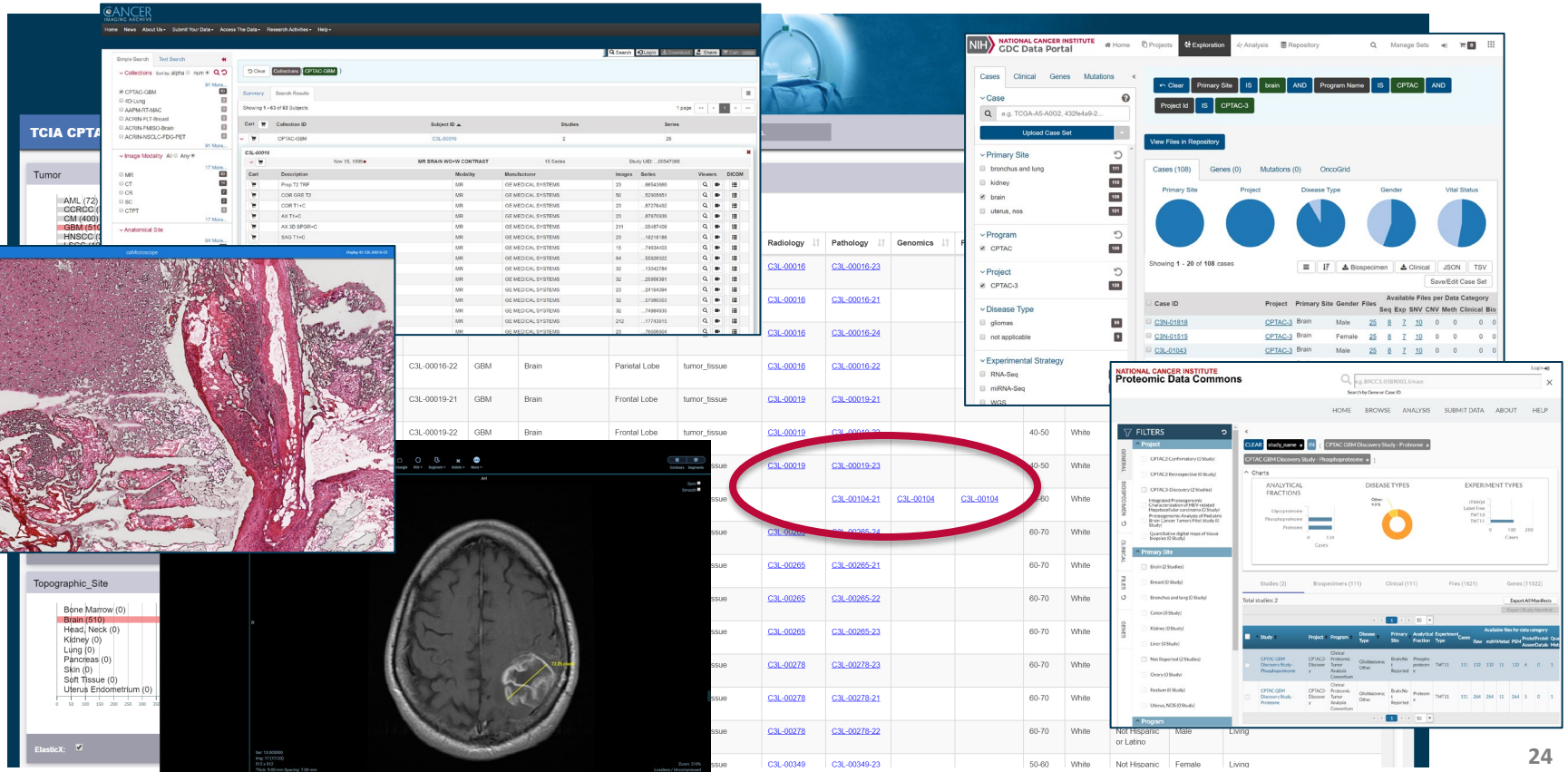
Data Access

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 t download a subset of its contents.

Data Type	Download all or Query/Filter
Images (DICOM, 39.4 GB)	Download Search
Tissue Slide Images (SVS, 87 GB)	Download Search
Clinical Data API (JSON - more info)	Download
Discovery Study Proteomics/Clinical Data (external)	• CPTAC Data Portal (Georgetown) • Proteomic Data Commons
Genomics/Clinical Data (External)	Genomic Data Commons



CPTAC Pathology Portal: Links to all data types



Cancer-specific links to other CPTAC Resources

CPTAC-GBM

Created by Tracy Nolan, last modified by natasha honomichl on Feb 14, 2020

Summary

This collection contains subjects from the National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium Glioblastoma Multiforme (CPTAC-GBI molecular basis of cancer through the application of large-scale proteome and genome analysis, or proteogenomics. Radiology and pathology images fr by The Cancer Imaging Archive to enable researchers to investigate cancer phenotypes which may correlate to corresponding proteomic, genomic and cl

CPTAC Phase 3 collects data from ten cancer types. In TCIA, imaging from each cancer type will be contained in its own TCIA Collection, with the collectic available on TCIA each quarter as it is collected. A summary of CPTAC Phase 3 imaging efforts can be found on the CPTAC Imaging Proteomics page.

Radiology imaging is collected from standard of care imaging performed on patients immediately before the pathological diagnosis, and from follow-up heterogeneous in terms of scanner modalities, manufacturers and acquisition protocols. Pathology imaging is collected as part of the CPTAC qualificatior

CPTAC Imaging Special Interest Group

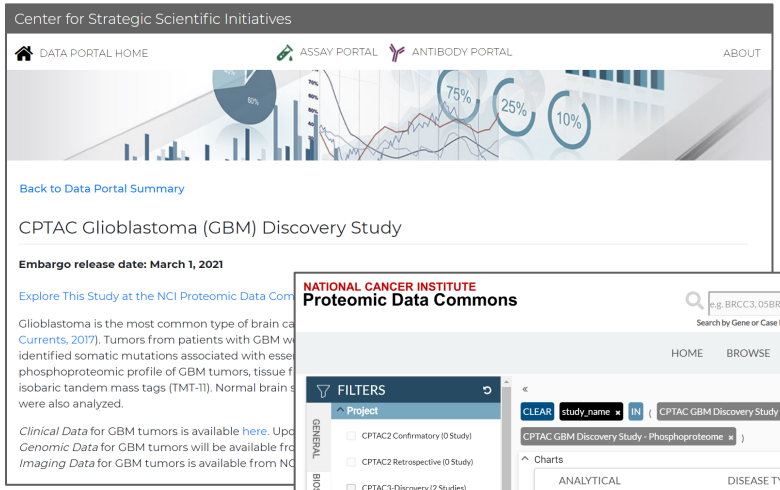
You can join the CPTAC Imaging Special Interest Group to be notified of webinars & data releases, collaborate on common data wrangling tasks and seek webinars such as slide decks and video recordings can be found on the CPTAC SIG Webinars page.

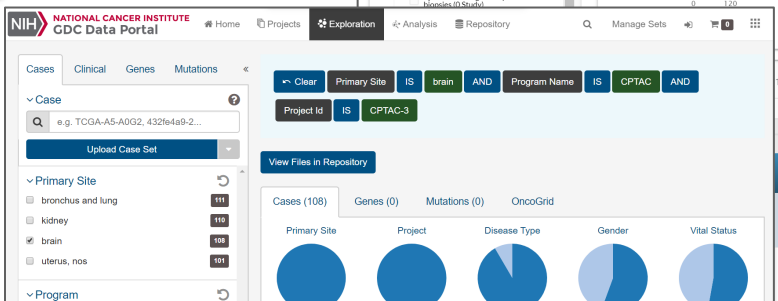
Data Access | Detailed Description | Citations & Data Usage Policy | Versions

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 t download a subset of its contents.

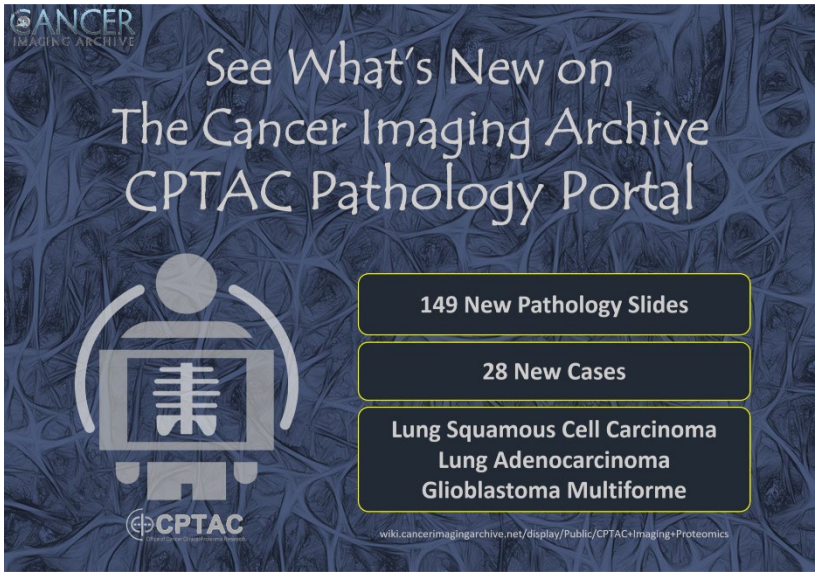
Data Type	Download all or Query/Filter
Images (DICOM, 39.4 GB)	Download Search
Tissue Slide Images (SVS, 87 GB)	Download Search
Clinical Data API (JSON - more info)	Download
Discovery Study Proteomics/Clinical Data (external)	• CPTAC Data Portal (Georgetown) • Proteomic Data Commons
Genomics/Clinical Data (External)	Genomic Data Commons



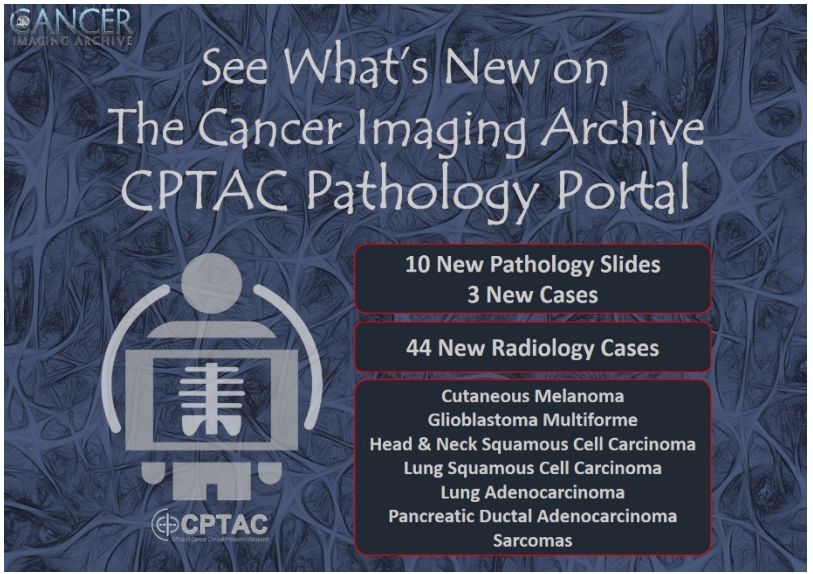


CPTAC Data Release Schedule

Monthly pathology updates



Quarterly radiology updates



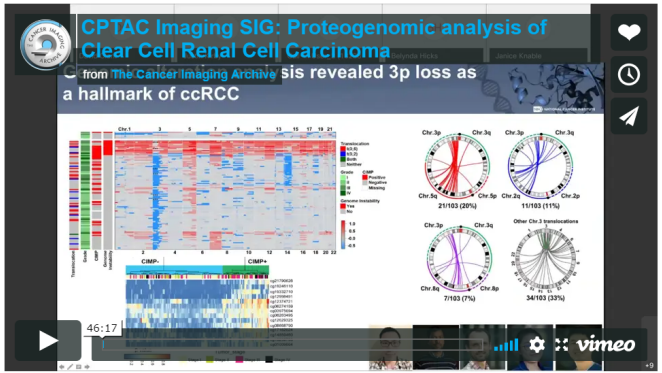
Get notified about CPTAC news as it becomes available!

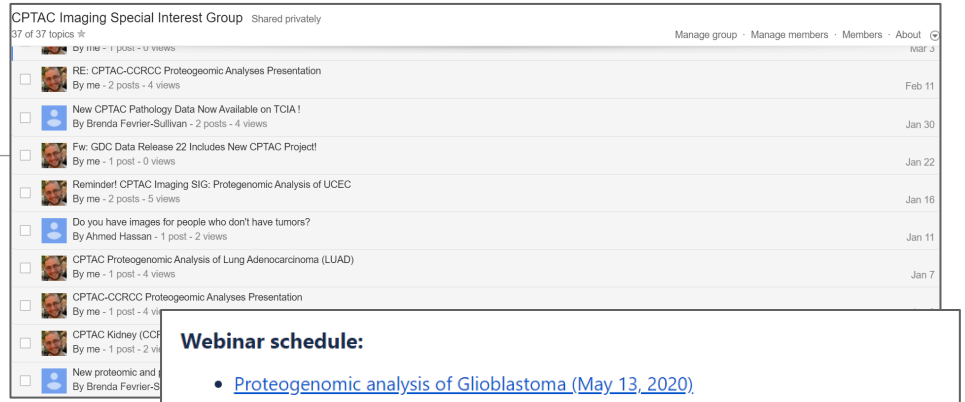
Click **here** to get email updates and follow **@NCI_CSSI** on Twitter

CPTAC Imaging Special Interest Group

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!

Proteogenomic analysis of Clear Cell Renal Cell Carcinoma (February 4, 2020)





Webinar schedule:

- Proteogenomic analysis of Glioblastoma (May 13, 2020)
- POSTPONED – Proteogenomic analysis of Lung Adenocarcinoma (March 3, 2020)
- Proteogenomic analysis of Clear Cell Renal Cell Carcinoma (February 4, 2020)
- Proteogenomic analysis of Uterine Corpus Endometrial Carcinoma (January 14, 2020)
- Imaging-Omic correlation studies utilizing CPTAC data (September 9, 2019)
- Accessing CPTAC data via Jupyter Notebooks (August 6, 2019)
- Program overview & data access tutorials (July 1, 2019)

Agenda & Slides

Dr. David Clarke presents the consortium's proteogenomic analyses of the CPTAC Clear Cell Renal Cell Carcinoma (CCRCC) cohort. This deep dive into the CCRCC genomic and proteomic datasets will help researchers better understand how these can be correlated with features derived from the imaging data. (**Download the slides**)

Follow TCIA on Social Media

Twitter: @TCIA_News

Facebook: The Cancer Imaging Archive (page)

LinkedIn: The Cancer Imaging Archive (group)

Acknowledgements

Fredrick National Laboratory for Cancer Research
John Freymann
Justin Kirby
Brenda Fevrier-Sullivan
Pam Angelus
Luis Cordeiro
Michelle Tacconelli

NBIA Team (NCI/FNLCR/Ellumen)
Ulli Wagner
Scott Gustafson
Qinyan Pan
Russ Rieling
Carolyn Klinger
Martin Lerner
Tin Tran
Keyvan Farahani
Ed Helton

University of Arkansas
Dr. Fred Prior
Kirk Smith
Lawrence Tarbox
Dwayne Dobbins
Jeff Tobler
Sonya Utecht
Julie Frund
Bill Bennett
Tracy Nolan
Quasar Jarosz
Brittney Camp
Erica Bilello
Claren Freeman
Natasha Honomichl
Geri Blake
Diana Stockton
Erica Bilello
Robert Brown

Emory University
Dr. Ashish Sharma
Ryan Birmingham

Contractors
John Perry
Carl Jaffe
Betty Levine

Imaging and Radiation Oncology Core
Dr. Michael Knopp
Connor Nealer
Rich Jacko
Jun Zhang