



Data Sharing via the NCI Proteomic Data Commons – Live Demo

CPTAC Scientific Symposium, September 2025



Ratna Rajesh Thangudu, Ph.D.

09/30/2025



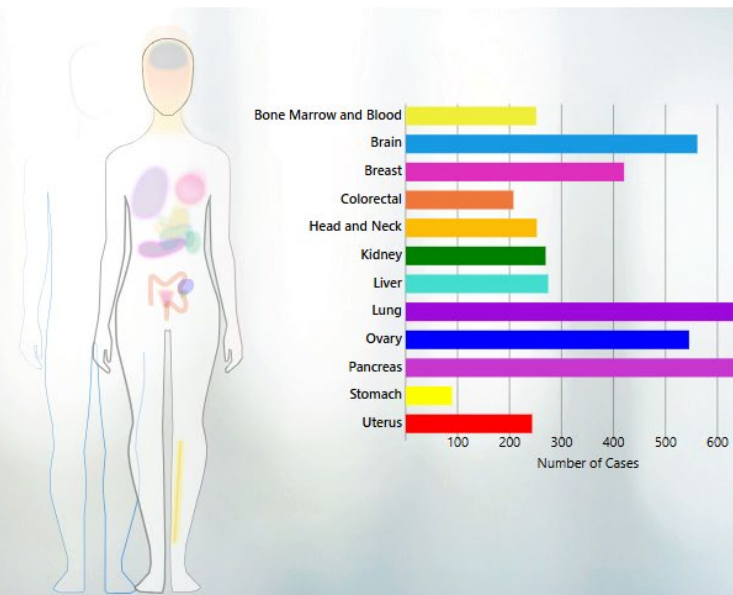
Agenda

- PDC Overview
- Harmonization and Standards
- Integration with CRDC
- Recent Enhancements
- Live Demo

Discover the Proteomic Data Commons (PDC)

Open access to harmonized cancer proteomic data, with visual connections and cross-references to genomic and imaging resources, helping researchers make multi-omic discoveries.

[Explore Data](#)



National Cancer Institute
at the National Institutes of Health

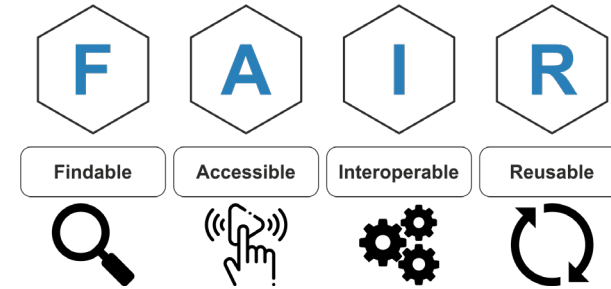


Building Blocks: FAIR & Harmonization

- Adherence to FAIR Guidelines
- Common Data Models & Standardization
- Harmonization Pipelines
- Sophisticated Tools for Data Exploration
- Analysis tools for protein/gene expression clustering, building cohorts, and survival analysis.
- APIs for programmatic access



Building Blocks: FAIR & Harmonization

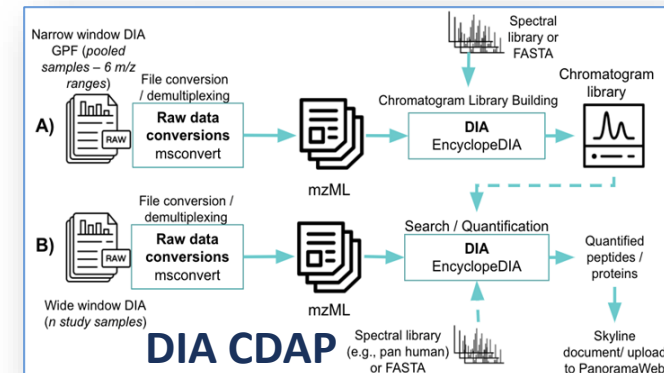
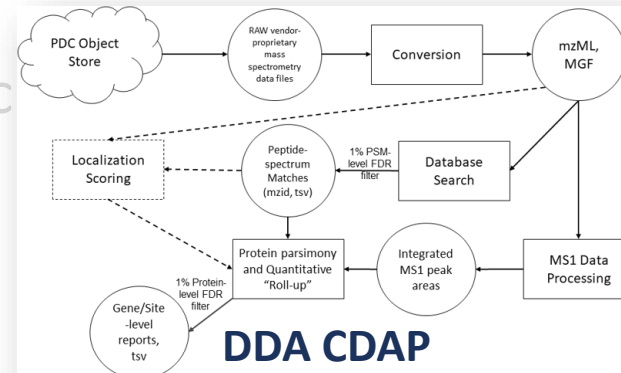
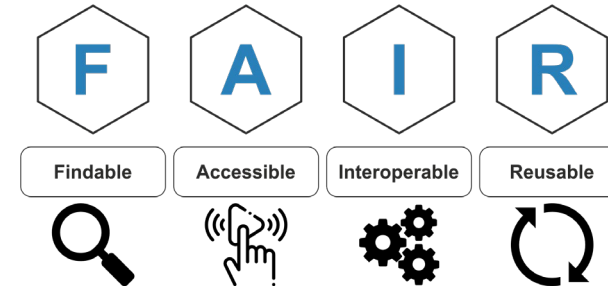


- Adherence to FAIR Guidelines
- Common Data Models & Standardization
- Harmonization Pipelines
- Sophisticated Tools for Data Exploration
- Analysis tools for protein/gene expression clustering, building cohorts, and survival analysis.
- APIs for programmatic access



Building Blocks: FAIR & Harmonization

- Adherence to FAIR Guidelines
- Common Data Models & Standardization
- Harmonization Pipelines
- Sophisticated Tools for Data Exploration
- Analysis tools for protein/gene expression clustering, building cohorts, and survival analysis.
- APIs for programmatic

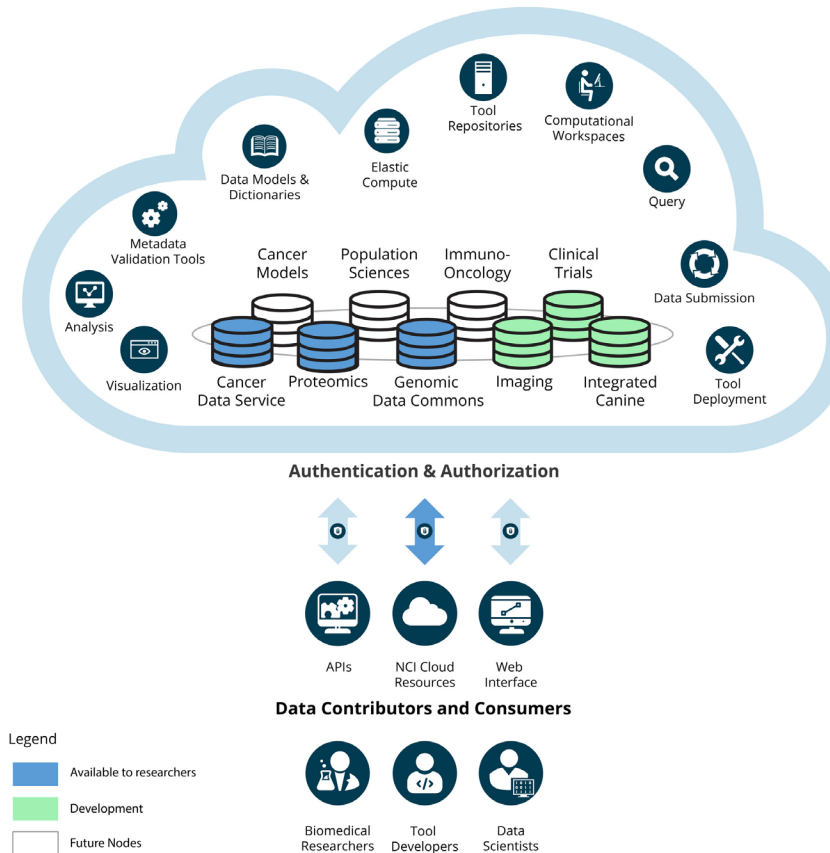


PDC in the CRDC Ecosystem

Why Specialized Resources Are Needed:

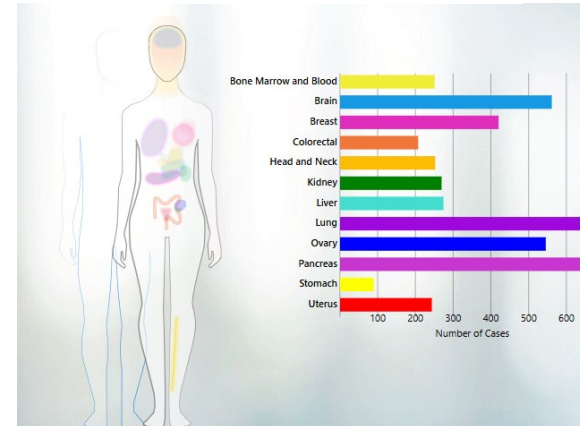
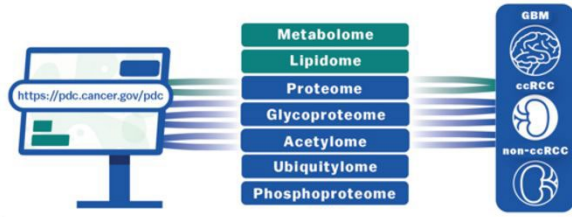
- Handling **complex, large-scale multi-omics** data.
- Ensuring **data standardization** and **harmonization** across studies.
- Providing access to cutting-edge **analytical tools** and **computational infrastructure**.
- Making data **FAIR** (Findable, Accessible, Interoperable, Reusable).
- Supporting the research community in reusing data to drive **scientific breakthroughs**.

NCI Cancer Research Data Commons (CRDC)



Proteomic Data Commons Now Supports Metabolomics and Lipidomics Data:

New Data From Three CPTAC Studies Released



<https://pdc.cancer.gov>

191
Studies

61 TB
Data Volume

178,833
Data Files

5,047
Cases

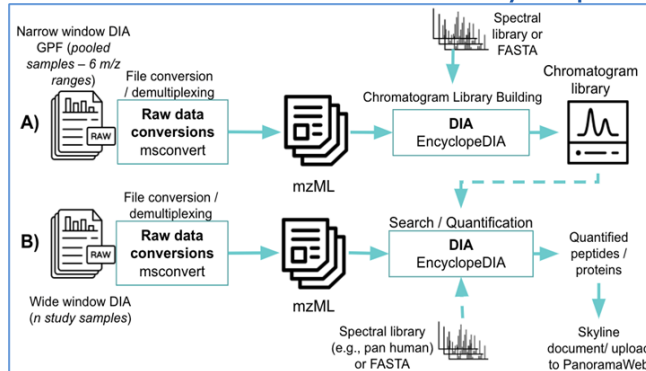
36
Disease Types

16,860
Proteins

Multi-Omics Availability

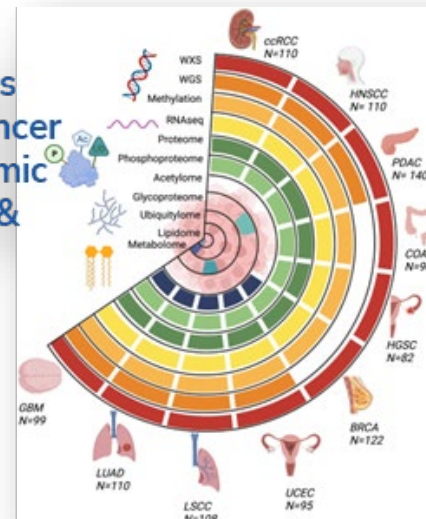
Genomics	Imaging	Lipidome	Metabolome
✓	✓	✓	✓
✓	✓	✓	✓
✓	✓	✓	✓
✓	✓	✓	✓

DIA Data Analysis Pipeline

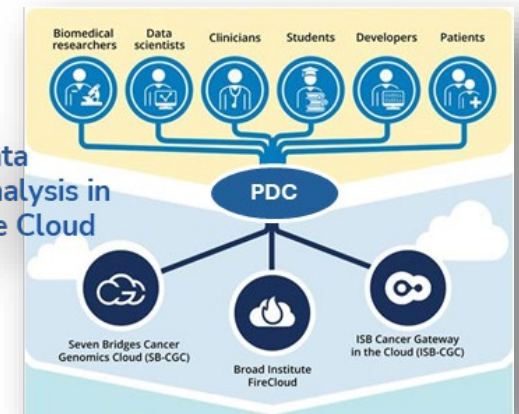


<https://nf-teirex-dia.readthedocs.io/en/latest/overview.html>

CPTAC's Pan-Cancer Multi-omic Papers & Data



Data Analysis in the Cloud



Live Demo: What Can We Ask in PDC?



- Which GBM studies have multi-omic data available beyond proteomics?
- Can I find breast cancer cases that are Stage II?
- Show me lung cancer cases that have undergone chemotherapy or radiation.
- How do I track a specific case across multiple studies and programs?
- How can I visualize protein quantitation data without downloading files?
- Is my gene of interest present in PDC studies?



Expanding CPTAC Data Sharing



<https://aacrjournals.org/cancerres/issue/84/9>

RESEARCH ARTICLE | SEPTEMBER 20 2024

NCI's Proteomic Data Commons: A Cloud-Based Proteomics Repository Empowering Comprehensive Cancer Analysis through Cross-Referencing with Genomic and Imaging Data



Ratna R. Thangudu ; Michael Holck ; Deepak Singhal ; Alexander Piloizzi ; Nathan Edwards ; Paul A. Rudnick ; Marcin J. Domagalski ; Padmini Chilappagari ; Lei Ma ; Yi Xin ; Toan Le ; Kristen Nyce ; Rekha Chaudhary ; Karen A. Ketchum ; Aaron Maurais ; Brian Connolly ; Michael Riffle ; Matthew C. Chambers ; Brendan MacLean ; Michael J. MacCoss ; Peter B. McGarvey ; Anand Basu ; John Otridge ; Esmeralda Casas-Silva ; Sudha Venkatachari ; Henry Rodriguez ; Xu Zhang

Check for updates

+ Author & Article Information

Cancer Research Communications (2024) 4 (9): 2480–2488.

<https://doi.org/10.1158/2767-9764.CRC-24-0243> [Article history](#)

AACR American Association
for Cancer Research



CANCER
RESEARCH
COMMUNICATIONS

<https://doi.org/10.1158/2767-9764.CRC-24-0243>



Resources

CPTAC Program	https://proteomics.cancer.gov	cancer.proteomics@mail.nih.gov
Proteomic Data commons	https://pdc.cancer.gov	pdhelpdesk@mail.nih.gov
Genomic Data Commons	https://gdc.cancer.gov	support@nci-gdc.datacommons.io
Imaging Data Commons	https://portal.imaging.datacommons.cancer.gov	support@canceridc.dev
The Cancer Imaging Archive	https://www.cancerimagingarchive.net	help@cancerimagingarchive.net
Assay Portal	https://assays.cancer.gov/	cancer.proteomics@mail.nih.gov
Antibody Portal	https://antibodies.cancer.gov/	cancer.proteomics@mail.nih.gov



Acknowledgements

PDC

Michael Holck
Alexander Pillozzi
Toan Le
Kristen Nyce
Rekha Chaudary

SME

Michael MacCoss
Nathan Edwards
Paul Rudnick

CPTAC DCC

Karen Ketchum
Yin Lu
Shuang Cai

NCI

OCCPR

Henry Rodriguez
Xu Zhang
Mehdi Mesri
Eunkyung An
Tara Hiltke
Jasmin Bavarva

CRDC

Emi Casas-Silva
CRDC Stakeholders
CBIIT Leadership

Leidos

Sudha Venkatachari
John Otridge

CDR

Mathangi Thiagarajan

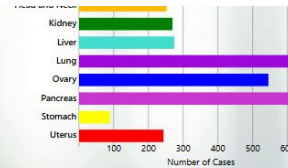
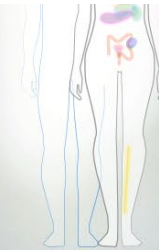
CPTAC Consortium

Data Donors



Visual connections and cross-references to genome and imaging resources, helping researchers make multi-omic discoveries.

[Explore Data](#)



191
Studies



61 TB
Data Volume



178,833
Data Files



5,047
Cases



36
Disease Types



16,860
Proteins

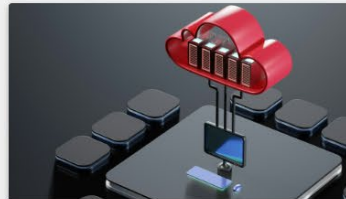
The PDC: A Core Resource of the NCI Cancer Research Data Commons (CRDC)

Offering advanced analytical resources and harmonized proteomic datasets for cancer research, and enabling researchers to share data across the community.



[CPTAC Pan-Cancer Data](#)

Explore data, publications, and supplementary materials from CPTAC's pan-cancer analysis.



[Cloud Data Analysis](#)

Analyze PDC datasets in the cloud using scalable workflows and integrated tools.



[How to Submit Data](#)

Submit your proteomics data to support cancer research and FAIR sharing.

Scientific Updates and Resources



Filter

General

Clinical

Genes

Sample

Case

Demographic

Diagnosis

Primary Site

Disease Type

☒ Glioblastoma (11 Studies)

☐ Acute Myeloid Leukemia (0 Study)

☐ Adenomas and Adenocarcinomas (0 Study)

☐ Breast Invasive Carcinoma (0 Study)

☐ Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma (0 Study)

☐ Cholangiocarcinoma (0 Study)

☐ Chromophobe Renal Cell Carcinoma (0 Study)

☐ Clear Cell Renal Cell Carcinoma (0 Study)

☐ Colon Adenocarcinoma (0 Study)

☐ Complex Epithelial Neoplasms (0 Study)

☐ Cystic, Mucinous and Serous Neoplasms (0 Study)

☐ Ductal and Lobular Neoplasms (0 Study)

☐ Early Onset Gastric Cancer (0 Study)

☐ Epithelial Neoplasms, NOS (0 Study)

☐ Gliomas (0 Study)

☐ Head and Neck Squamous Cell Carcinoma (0 Study)

☐ Hepatocellular Carcinoma (0 Study)

☐ Lung Adenocarcinoma (0 Study)

☐ Lung Squamous Cell Carcinoma (0 Study)

☐ Meningiomas (0 Study)

☐ Myelodysplastic Syndromes (0 Study)

☐ Neuroendocrine Neoplasms (0 Study)

☐ Non-Clear Cell Renal Cell Carcinoma (0 Study)

☐ Not Applicable (0 Study)

☐ Not Reported (0 Study)

☐ Oral Squamous Cell Carcinoma (0 Study)

☐ Other (0 Study)

CLEAR

program_name

IS

CPTAC

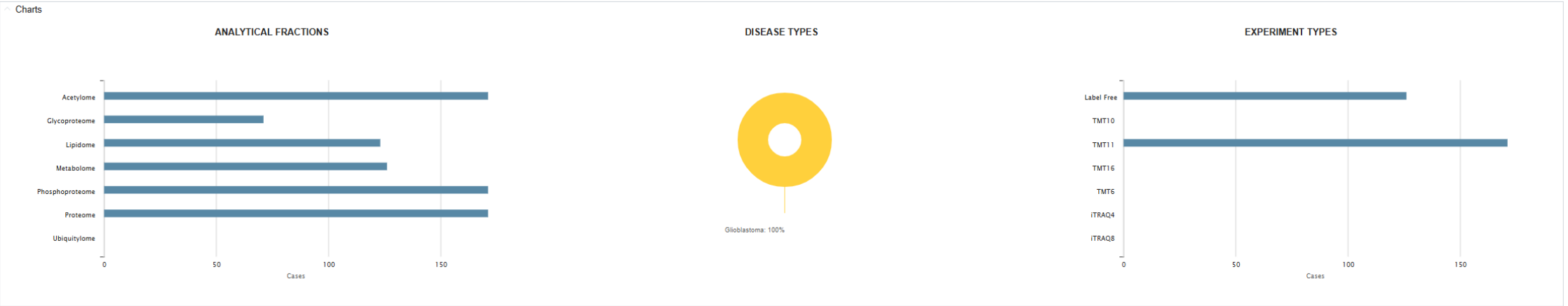
AND

disease_type

IS

Glioblastoma

Show Query URL



Studies (11)	Biospecimens (299)	Clinical (171)	Files (5066)	Genes (12010)
--------------	--------------------	----------------	--------------	---------------

Total studies: 11

Export Study Manifest

	PDC Study ID	Study	Program	Disease Type	Primary Site	Analytical Fraction	Experiment Type	Cases #	Multi-omics Availability			
									Genomics	Imaging	Lipidome	Metabolome
☐												
☒	PDC000553	CPTAC GBM Discovery Study - Lipidome	Clinical Proteomic Tumor Analysis Consortium	Glioblastoma; Other	Brain	Lipidome	Label Free	83	✓	✓	✓	✓
☒	PDC000552	CPTAC GBM Discovery Study - Metabolome	Clinical Proteomic Tumor Analysis Consortium	Glioblastoma; Other	Brain	Metabolome	Label Free	83	✓	✓	✓	✓
☒	PDC000547	CPTAC GBM Confirmatory Study - Lipidome	Clinical Proteomic Tumor Analysis Consortium	Breast Invasive Carcinoma; Glioblastoma; Gliomas; Lung Adenocarcinoma; Meningiomas; Other	Brain; Breast; Bronchus and lung; Not Reported	Lipidome	Label Free	82	✓	✓	✓	✓
☒	PDC000546	CPTAC GBM Confirmatory Study - Metabolome	Clinical Proteomic Tumor Analysis Consortium	Breast Invasive Carcinoma; Glioblastoma; Gliomas; Lung Adenocarcinoma; Meningiomas; Other	Brain; Breast; Bronchus and lung; Not Reported	Metabolome	Label Free	85	✓	✓	✓	✓
☒	PDC000454	CPTAC GBM Confirmatory Study - Glycoproteome	Clinical Proteomic Tumor Analysis Consortium	Breast Invasive Carcinoma; Complex Epithelial Neoplasms; Epithelial Neoplasms, NOS; Glioblastoma; Gliomas; Lung Adenocarcinoma; Meningiomas; Other; Skin Cutaneous Melanoma; Uterine Adenocarcinoma	Brain; Breast; Bronchus and lung; Colon; Not Reported; Other and unspecified urinary organs; Unknown; Uterus; NOS	Glycoproteome	TMT11	118	✓	✓	✓	✓
☒	PDC000450	CPTAC GBM Confirmatory Study - Acetylome	Clinical Proteomic Tumor Analysis Consortium	Breast Invasive Carcinoma; Complex Epithelial Neoplasms; Epithelial Neoplasms, NOS; Glioblastoma; Gliomas; Lung Adenocarcinoma; Meningiomas; Other; Skin Cutaneous Melanoma; Uterine Adenocarcinoma	Brain; Breast; Bronchus and lung; Colon; Not Reported; Other and unspecified urinary organs; Unknown; Uterus; NOS	Acetylome	TMT11	118	✓	✓	✓	✓

Filter

General

Clinical

Genes

Sample

Case

Demographic

Ethnicity

Race

Sex

Vital Status

Diagnosis

Primary Site

Disease Type

Age at Diagnosis

Ajcc Clinical Stage

Ajcc Pathologic Stage

Progression Or Recurrence

Tumor Grade

Morphology

Site of Resection or Biopsy

Parietal lobe (4 Studies)

Abdomen, NOS (0 Study)

Appendix (0 Study)

Ascending colon (0 Study)

Basal Ganglia (0 Study)

Basal Ganglia,Temporal Lobe (0 Study)

Basal Ganglia,Thalamus (0 Study)

Base of tongue, NOS (0 Study)

Blood (0 Study)

CLEAR

program_name

IS

CPTAC

AND

disease_type

IS

Glioblastoma

AND

race

IS

black or african american

AND

site_of_resection_or_biopsy

IS

Parietal lobe

Show Query URL

Charts

ANALYTICAL FRACTIONS

DISEASE TYPES

EXPERIMENT TYPES

Studies (4)

Biospecimens (1)

Clinical (1)

Files (2547)

Genes (11603)

Total studies: 4

1

10

	PDC Study ID	Study	Program	Disease Type	Primary Site	Analytical Fraction	Experiment Type	Cases #	Multi-omics Availability			
									Genomics	Imaging	Lipidome	Metabolome
☐	PDC000454	CPTAC GBM Confirmatory Study - Glycoproteome	Clinical Proteomic Tumor Analysis Consortium	Breast Invasive Carcinoma; Complex Epithelial Neoplasms; Epithelial Neoplasms, NOS; Glioblastoma; Gliomas; Lung Adenocarcinoma; Meningiomas; Other; Skin Cutaneous Melanoma; Uterine Adenocarcinoma	Brain; Breast; Bronchus and lung; Colon; Not Reported; Other and unspecified urinary organs; Unknown; Uterus, NOS	Glycoproteome	TMT11	118	✓	✓	✓	✓
☐	PDC000450	CPTAC GBM Confirmatory Study - Acetylome	Clinical Proteomic Tumor Analysis Consortium	Breast Invasive Carcinoma; Complex Epithelial Neoplasms; Epithelial Neoplasms, NOS; Glioblastoma; Gliomas; Lung Adenocarcinoma; Meningiomas; Other; Skin Cutaneous Melanoma; Uterine Adenocarcinoma	Brain; Breast; Bronchus and lung; Colon; Not Reported; Other and unspecified urinary organs; Unknown; Uterus, NOS	Acetylome	TMT11	118	✓	✓	✓	✓
☐	PDC000448	CPTAC GBM Confirmatory Study - Phosphoproteome	Clinical Proteomic Tumor Analysis Consortium	Breast Invasive Carcinoma; Complex Epithelial Neoplasms; Epithelial Neoplasms, NOS; Glioblastoma; Gliomas; Lung Adenocarcinoma; Meningiomas; Other; Skin Cutaneous Melanoma; Uterine Adenocarcinoma	Brain; Breast; Bronchus and lung; Colon; Not Reported; Other and unspecified urinary organs; Unknown; Uterus, NOS	Phosphoproteome	TMT11	118	✓	✓	✓	✓
					Brain; Breast;							

1

10

NIH NATIONAL CANCER INSTITUTE

STUDY SUMMARY: CPTAC GBM Confirmatory Study - Proteome

118

Cases

150

Aliquots

SUMMARY

PDC Study Identifier	PDC000446	Analytical Fraction	Proteome
Study ID	b25c2cea-4a49-4bc3-9f87-3cf5ee026865	Disease Types	Breast Invasive Carcinoma;Complex Epithelial Neoplasms;Epithelial Neoplasms, NOS;Glioblastoma;Gliomas;Lung Adenocarcinoma;Meningiomas;Other;Skin Cutaneous Melanoma;Uterine Adenocarcinoma
Study Name	CPTAC GBM Confirmatory Study - Proteome	Project ID	CPTAC3 Discovery and Confirmatory
Experiment Type	TMT11		

MULTI-OMIC AVAILABILITY

Genomics	Available - View Data
Imaging	Available - View Data
Lipidome	Available - View Data ▼
Metabolome	Available - View Data ▼

Description 2

Protocol (1) 2

Experimental Design 2

Clinical 2

Biospecimens 2

Workflow 2

DUA 2

To delineate the multi-scale regulatory interactions governing glioblastoma (GBM) development and evolution (recurrence and clinical outcomes), we characterized 228 human GBM tumors (including 28 at recurrence) using 15 proteogenomic and metabolomic platforms. Comparative analyses of paired primary-recurrent GBMs showed increased clonal diversity in recurrent tumors as a function of elapsed time since primary diagnosis and treatment-induced mutational and signaling signatures at recurrence. Additionally, across our entire cohort, independent analyses of proteomic and metabolomic changes associated with the trans effects of PTEN and TERT promoter mutations, when considered independently, showed a high correlation between the downstream effects of the two drivers. Glycoproteomic and phosphoproteomic analyses revealed crosstalk of dual glycosylation (N352 and N603) and phosphorylation (Y316) on EGFR, but only observed in the context of EGFR amplification or activating mutations. IDH1 mutation was associated with activated RTK signaling and decreased hypoxia pathway activities, which was concordant with epigenetic and metabolic profiles. Protein-protein interaction and kinase/phosphatase-substrate analyses uncovered detailed signaling flowing from upstream drivers (e.g., EGFR, PDGFRA, and IDH) through a PTPN11 network hub to downstream effectors, including GAB1, IRS1, MAP3K5, and PTK2B. We also report recurrent mutations in PTPN11 and its downstream consequences. In summary, this study presents new biological insights regarding treatment impact on tumor evolution and clinical outcomes, shared downstream consequences of independent drivers, and the potential importance of members of PTPN11 signaling circuitry across high-grade gliomas.

Note: The samples listed below are metastatic samples originating from organs including lung, uterus, colon, breast or urethra which metastasized to the brain and were subsequently resected from the brain.

7316UP-483, 7316UP-99, 7316UP-1273, 7316UP-743, 7316UP-219, 7316UP-485, 7316UP-393, 7316UP-553, 7316UP-1302, 7316UP-1405, 7316UP-1614, 7316UP-1883

Common Data Analysis Pipeline (PDC Harmonization) data 2

Data Category	Files (n=1447)
Peptide Spectral Matches (Open Standard)	360
Peptide Spectral Matches (Text)	360
Processed Mass Spectra (Open Standard)	360
Protein Assembly (Text)	5
Quality Metrics (Text)	1
Quality Metrics (Web)	1
Raw Mass Spectra (Proprietary)	360

External References

Genomic Data Commons	CPTAC-3
Imaging Data Commons	CPTAC-GBM
Database of Genotypes and Phenotypes	phs001287
The Cancer Imaging Archive	CPTAC-GBM

Related PDC studies

CPTAC GBM Confirmatory Study - Glycoproteome	PDC000454
CPTAC GBM Confirmatory Study - Metabolome	PDC000546
CPTAC GBM Confirmatory Study - Lipidome	PDC000547
CPTAC GBM Confirmatory Study - CompRef Proteome	PDC000447
CPTAC GBM Confirmatory Study - Phosphoproteome	PDC000448
CPTAC GBM Confirmatory Study - CompRef Phosphoproteome	PDC000449
CPTAC GBM Confirmatory Study - Acetylome	PDC000450
CPTAC GBM Confirmatory Study - CompRef Acetylome	PDC000451

Supplementary data 2

Data Category	Files (n=1)
Publication Supplementary Material (Archive)	1



Explore protein quantitation from PDC Common Data Analysis pipeline (CDAP) through heatmaps

PUBLICATIONS



Liu J., Cao S., Imbach KJ., Gritsenko MA., et al. (2024). Multi-scale signaling and tumor evolution in high-grade gliomas. Cancer Cell 42(7) 1217-1238.

118
Cases

150
Aliquots

SUMMARY

PDC Study Identifier

PDC000446

Study ID

b25c2cea-4a49-4bc3-9f87-3cf5ee026865

Study Name

CPTAC GBM Confirmatory Study - Proteome

Experiment Type

TMT11

Analytical Fraction

Proteome

Disease Types

Breast Invasive Carcinoma;Complex Epithelial Neoplasms;Epithelial Neoplasms, NOS;Glioblastoma;Gliomas;Lung Adenocarcinoma;Meningiomas;Other;Skin Cutaneous Melanoma;Uterine Adenocarcinoma

Project ID

CPTAC3 Discovery and Confirmatory

MULTI-OMIC AVAILABILITY

Genomics

Available - View Data

Imaging

Available - View Data

Lipidome

Available - View Data

Metabolome

Available - View Data

Description

Protocol (1)

Experimental Design

Clinical

Biospecimens

Workflow

DUA

Total records: 118

Export CSV TSV											
Case Submitter ID	Genomic Data Resource	Imaging Data Resource	Ethnicity	Sex	Race	Morphology	Primary Diagnosis	Site of Resection or Biopsy	Tissue or Organ of Origin	Tumor Grade	Tumor Stage
C247230	17b917b5-f10d-49eb-8dcf-1f517a838d30		not reported	male	not reported	9380/3	Glioma, NOS	Cerebrum	Cerebrum	G4	Not Reported
C3N-04686	06055207-d5ba-4b0e-a6d7-81bcf2bbd149	TCIA.2018.3RJE41Q1	not reported	male	white	9440/3	Glioblastoma	Frontal lobe	Brain, NOS	Not Reported	Not Reported
C3N-03484	a8e64aea-fb94-4c83-b96a-126a718ce3dd	TCIA.2018.3RJE41Q1	not reported	female	asian	9440/3	Glioblastoma	Frontal lobe	Brain, NOS	Not Reported	Not Reported
C1245129	6cd85f15-79df-45b5-acda-b938eb35fccb		not reported	male	not reported	9440/3	Glioblastoma	Temporal lobe	Temporal lobe	G4	Not Reported
C3N-01191	daf423d7-81d9-41b9-9d1c-5b06bb40d0c0	TCIA.2018.3RJE41Q1	not reported	female	white	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported
C3N-03452	f54ddbc9-5251-4db0-ba83-0dae7d2ce40c	TCIA.2018.3RJE41Q1	not reported	male	white	9440/3	Glioblastoma	Frontal lobe	Brain, NOS	Not Reported	Not Reported
C3L-04213	029e2f23-a4a4-4529-a428-336025bcc059	TCIA.2018.3RJE41Q1	not reported	female	white	9440/3	Glioblastoma	Occipital lobe	Brain, NOS	Not Reported	Not Reported
C464694	df5cca46-9ea2-4105-b298-901a9329614f		not reported	female	not reported	9440/3	Glioblastoma	Frontal lobe	Frontal lobe	G4	Not Reported
ref			not reported	not reported	not reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported	Not Reported
C3N-03440	54d05403-0875-4207-9b50-40c876c31a57	TCIA.2018.3RJE41Q1	not reported	male	white	9440/3	Glioblastoma	Temporal lobe	Brain, NOS	Not Reported	Not Reported

Common Data Analysis Pipeline (PDC Harmonization) data	
Data Category	Files (n=1447)
Peptide Spectral Matches (Open Standard)	360
Peptide Spectral Matches (Text)	360
Processed Mass Spectra (Open Standard)	360
Protein Assembly (Text)	5
Quality Metrics (Text)	1
Quality Metrics (Web)	1
Raw Mass Spectra (Proprietary)	360

External References	
Genomic Data Commons	CPTAC-3
Imaging Data Commons	CPTAC-GBM
Database of Genotypes and Phenotypes	phs001287
The Cancer Imaging Archive	CPTAC-GBM

Related PDC studies	
CPTAC GBM Confirmatory Study - Glycoproteome	PDC000454
CPTAC GBM Confirmatory Study - Metabolome	PDC000546
CPTAC GBM Confirmatory Study - Lipidome	PDC000547
CPTAC GBM Confirmatory Study - CompRef Proteome	PDC000447

Supplementary data	
Data Category	Files (n=1)
Publication Supplementary Material (Archive)	1



Explore protein quantitation from PDC Common Data Analysis pipeline (CDAP) through heatmaps

PUBLICATIONS

 Liu J., Cao S., Imbach KJ., Gritsenko MA., et al. (2024). Multi-scale signaling and tumor evolution in high-grade gliomas. Cancer Cell 42(7) 1217-1238.

Filter

General

Clinical

Demographic

Diagnosis

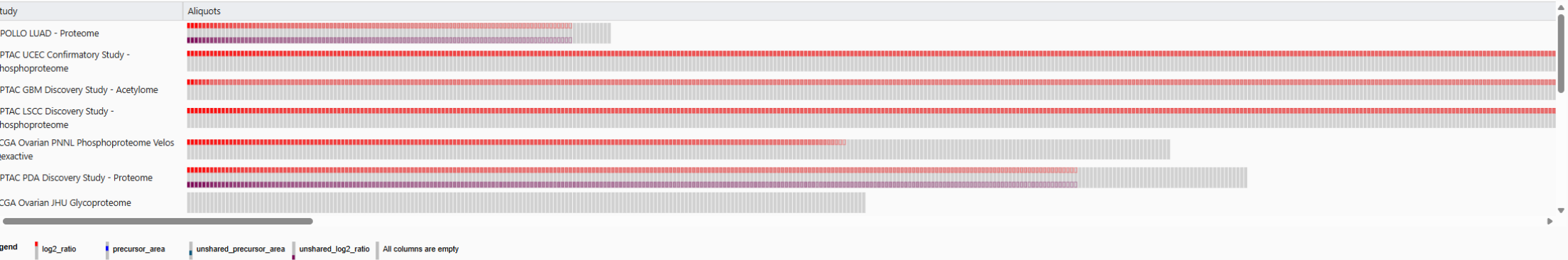
Treatment

Exposure

GENE: EGFR

NCBI Gene ID	1956	Chromosome	7
Authority	HGNC:3236	Locus	7p11.2
Description	epidermal growth factor receptor	Proteins	XP_047275908;XP_047275909;XP_054213392;XP_054213393;NP_001333826;NP_001333827;NP_001333828;NP_001333829;NP_001333870;NP_005219;NP_958439;NP_958440;NP_958441;P00533;C9JVS6;Q504U8;A0A8V8TPW8;ENSP00000345973;ENSP00000275493;ENSP00000415559;ENSP00000342376;ENSP00000413843;ENSP00000413354;ENSP00000413355
Organism	Homo sapiens	Aliases	ERRP;ERBB1;EGFR;ERBB
Assays	CPTAC-802;CPTAC-1043;CPTAC-1781		

Detected Post-translational Modifications			Total PTM sites: 175
PTM Type	Site	Peptide	
phospho	NP_001333828.1:t648	LLQERELVEPLtPSGEAPNQALLR;ELVEPLtPSGEAPNQALLR	
phospho	NP_001333828.1:t648s650	ELVEPLtPsGEAPNQALLR	
phospho	NP_001333828.1:s650	ELVEPLTPsGEAPNQALLR	
phospho	NP_001333828.1:s675	VLGsGAFGTVYK	
phospho	NP_001333828.1:t680	VLGSAGFGtVYK	



Studies in Which the Gene Product Was Detected								Total Studies: 58
Export Study Manifest								
	PDC Study ID	Study	Experiment Type	Spectral Counts	Distinct Peptides	Unshared Peptides	No of Aliquots	No of Plexes
☐	PDC000434	APOLLO LUAD - Proteome	TMT11	475	35	33	101	10
☐	PDC000441	CPTAC UCEC Confirmatory Study - Phosphoproteome	TMT11	206	13	12	161	16

CASE SUMMARY: C3L-00908

CASE DETAILS

Case	C3L-00908	Disease Type	Clear Cell Renal Cell Carcinoma
Case ID	ee6b0cb5-1fb8-11e9-b7f8-0a80fada099c	Program	Clinical Proteomic Tumor Analysis Consortium
Project	CPTAC3 Discovery and Confirmatory	Primary Site	Kidney

DEMOGRAPHY

Ethnicity	not hispanic or latino	Sex	female
Demographic Submitter Id	C3L-00908-DM	Race	white
Cause Of Death	Not Available	Days To Birth	-22179

Show more

DIAGNOSIS (1)

EXPOSURE (1)

FOLLOW UP (1)

TREATMENT (0)

FILE COUNTS BY EXPERIMENTAL STRATEGY

Aquisition Type	Study	Experimental Strategy	Files
DDA	CPTAC CCRCC Discovery Study - Intact Glycoproteome	TMT10	600
DDA	CPTAC CCRCC Discovery Study - Phosphoproteome	TMT10	1217
DDA	CPTAC CCRCC Discovery Study - Proteome	TMT10	2320
DIA	CPTAC CCRCC Confirmatory Study - DIA Intact Glycoproteome	Label Free	209
DIA	CPTAC CCRCC Confirmatory Study - DIA Phosphoproteome	Label Free	208
DIA	CPTAC CCRCC Confirmatory Study - DIA Proteome	Label Free	211
DIA	CPTAC CCRCC Discovery Study - DIA Proteome	Label Free	208

FILE COUNTS BY DATA CATEGORY

Data Category	Files
Publication Supplementary Material (Archive)	21
Other Metadata (Document)	23
Supplementary Data (Document)	2
Peptide Spectral Matches (Open Standard)	874
Processed Mass Spectra (Open Standard)	1173
Raw Mass Spectra (Proprietary)	1952
Spectral Library (Proprietary)	37
Alternate Processing Pipeline (Text)	4
Peptide Spectral Matches (Text)	874
Protein Assembly (Text)	9
Quality Metrics (Text)	2
Quality Metrics (Web)	2

BIOSPECIMEN

SAMPLES (3)

ALIQUOTS (3)

Sample Submitter ID	C3L-00908-06
Sample UUID	dc2b8047-204c-11e9-b7f8-0a80fada099c
Sample Type	Solid Tissue Normal
Status	Qualified
Pool	No
Sample is reference	No

GENOMIC AND IMAGING DATA RESOURCES

External References

Genomic Data Commons	763e0702-8379-4b5e-95d1-a84f412c51e7
The Cancer Imaging Archive	CPTAC-CCRCC



Work completed this month

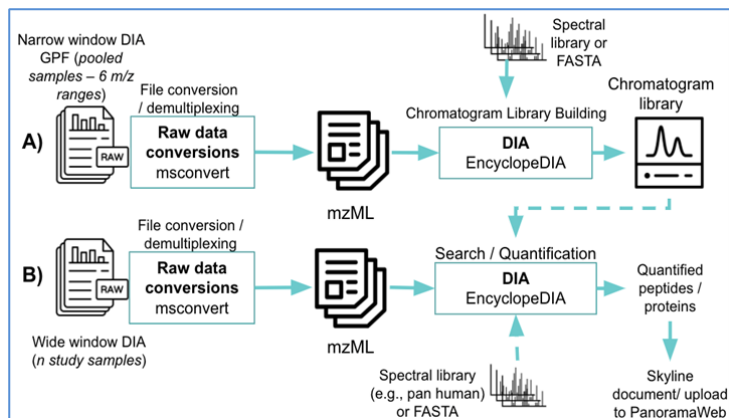
DIA Pipeline

- Successfully executed the pipeline on PDC infrastructure, action items identified

- EncyclopeDIA was tested on the PDA BioText study on PDC infrastructure
- Bruker .d support – currently in testing
- Syncing protein FASTA file across DDA and DIA
- Access data from PDC S3 to reduce egress - complete
- Ensuring appropriate identifiers are used

- QC and benchmarking evaluation in progress

- Release planned – DB 4.14



Search docs

Skyline DIA Nextflow Documentation

Workflow Overview

DIA-NN workflow:
EncyclopeDIA workflow:
How to Run
Workflow Components
How to Install the Workflow
How to Run the Workflow
Workflow Parameters
Output & Results
How to Setup and Configure AWS Batch

Workflow Overview

These documents describe a standardized Nextflow workflow for processing [DIA mass spectrometry data to quantify peptides and proteins](#). The source code for the workflow can be found at: <https://github.com/mrifle/nf-skyline-dia-ms>.

Multiple specific workflows may be run with this Nextflow workflow. Note that in all cases, the workflow can automatically generate requested reports from the Skyline document and can automatically upload and (optionally) import the Skyline document into PanoramaWeb and ProteomeXchange.

DIA-NN workflow:

The workflow will quantify peptides and proteins using user-supplied DIA RAW (or mzML) files, FASTA file, and spectral library (optional). If the user does not specify a spectral library, DIA-NN will be run in "library-free" mode, where it will create its own library using AI. Finally the workflow will generate a Skyline document using the quantified peptides and proteins.

EncyclopeDIA workflow:

This workflow is summarized in the following article:

[Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry](#). Searle BC, Pino LK, Egerton JD, Ting YS, Lawrence RT, MacLean BX, Villén J, MacCoss MJ. *Nat Commun*. 2018 Dec 3;9(1):5128. (<https://pubmed.ncbi.nlm.nih.gov/30510204/>)

The workflow will quantify peptides and proteins using user-supplied DIA RAW (or mzML) files, FASTA file, and spectral library. If the experimental design includes generation of a chromatogram library using narrow window DIA data, the workflow will first generate the chromatogram library (Figure 1A) and use that as input to the next phase (Figure 1B) to quantify peptides and proteins. If the experimental design does not include this, the user-supplied spectral library is used as input for

<https://nf-teirex-dia.readthedocs.io/en/latest/overview.html>

<https://github.com/mrifle/nf-skyline-dia-ms>