



---

# NCI Proteomic Data Commons Scientific Committee Meeting

---

09/28/2020, 5:00 – 6:00 PM ET

NATIONAL CANCER INSTITUTE CONTRACT: 20X042F01

## The Team

---

- **ESAC Inc., Rockville, MD**  
Anand Basu, Program Manager  
Ratna (Rajesh) Thangudu, Project Lead  
Michael Holck, Technical Lead
- **Leidos Biomedical., Fredrick, MD**  
John Otridge, Program Manager  
Sudha Venkatachari, Project Manager

- **University of Washington, Seattle, WA**  
Michael J MacCoss (Chair, Scientific Committee)
- **Georgetown University, Washington, DC**  
Nathan Edwards (SME Data Analysis)
- **Spectragen Informatics LLC, WA**  
Paul Rudnick (SME Data Analysis)
- **National Cancer Institute, Bethesda, MD**  
Henry Rodriguez  
Erika Kim

## Agenda

---

- PDC overview & goals - Mike MacCoss (15 min)
  - SC goals and responsibilities
  - Expectations - what specific feedback/inputs we seek from the SC
- CRDC Overview – Allen Dearry/Erika Kim (10 min)
- Current data and where we are today, Interoperability, System Architecture - Rajesh Thangudu (10 min)
- Data harmonization, APIs, CDAP and other pipelines - Paul Rudnick (10 min)
- Open discussion (15min)

# The PDC Scientific Committee

---

Member
Alexey Nesvizhskii, Ph.D. University of Michigan
Amanda Paulovich, M.D., Ph.D. Fred Hutchinson Cancer Research Institute
Bing Zhang, Ph.D. Baylor College of Medicine
Eric Deutsch, Ph.D. Institute for Systems Biology
Oliver Bogler, Ph.D. CCT/NCI
Sam Payne, Ph.D. Brigham Young University

## PDC -- High Level Goals

---

- Unsilo mass spectrometry data. Bring data into a common location that satisfy Findability, Accessibility, and Reusability
- Move from a situation where people move data to local tools to where people move their tools to the data.
- Shift from a 'data graveyard model' to a 'data workspace model'
- Make it feasible for pipelines to be released with data during publication to improve reproducibility
- Improve meta-data annotations. Ensure data is annotated well using common vocabularies but that the process is non-onerous.

## PDC Scientific Committee Responsibilities

---

- Provide constructive criticism on the work we have performed to date.
- Help us define the scope of the PDC.
- Help us prioritize our short-term and long-term to-do list.
- Provide comments about the UI/UX.
- Identify key collaborators to best demonstrate features and to contribute valuable cancer proteomics datasets
- Identify critical tools and workflows to integrate with the PDC to facilitate data reanalysis.
- Identify specific features that can make proteogenomic data accessible to cancer biologists and clinicians



## PDC Scientific Committee Commitment

---

- A 1-hour scientific committee meeting per quarter
- Occasional individual calls/emails regarding a specific issue we need input – up to 1-hour per week max.



# Overview of CRDC

---

**Dr Allen Dearry**



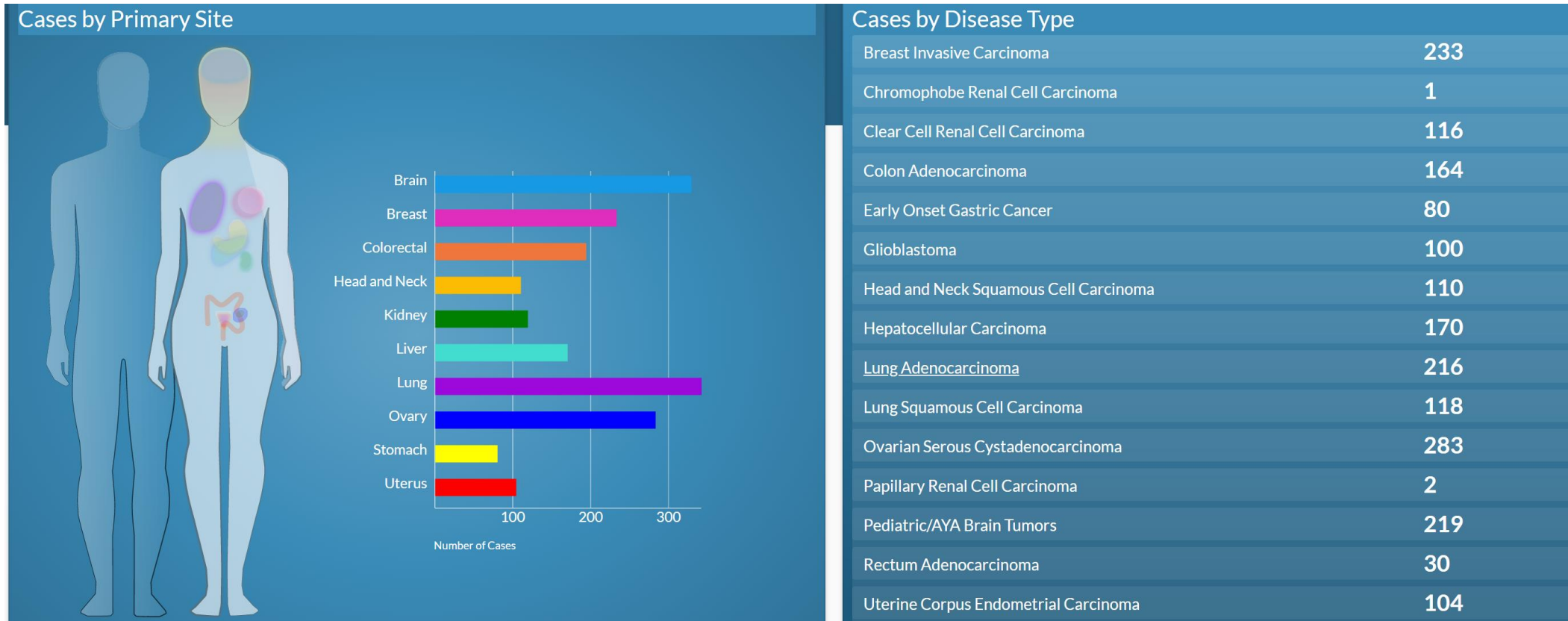


# Current data and where we are today, Interoperability

---

R Rajesh Thangudu

## PDC- Current status



## PDC by the numbers

---



### STUDIES

From large programs and also smaller labs.



### ACQUISITION TYPES

Data Dependent Acquisition  
Data Independent Acquisition.



### VOLUME

25 TB data including raw and processed data and supplementary information



### EXPERIMENTAL TYPES

Lable Free  
iTRAQ  
TMT



### ANALYTICAL FRACTIONS

Proteome, Phosphoproteome,  
Acetylome, Glycoproteome,  
Ubiquitylome

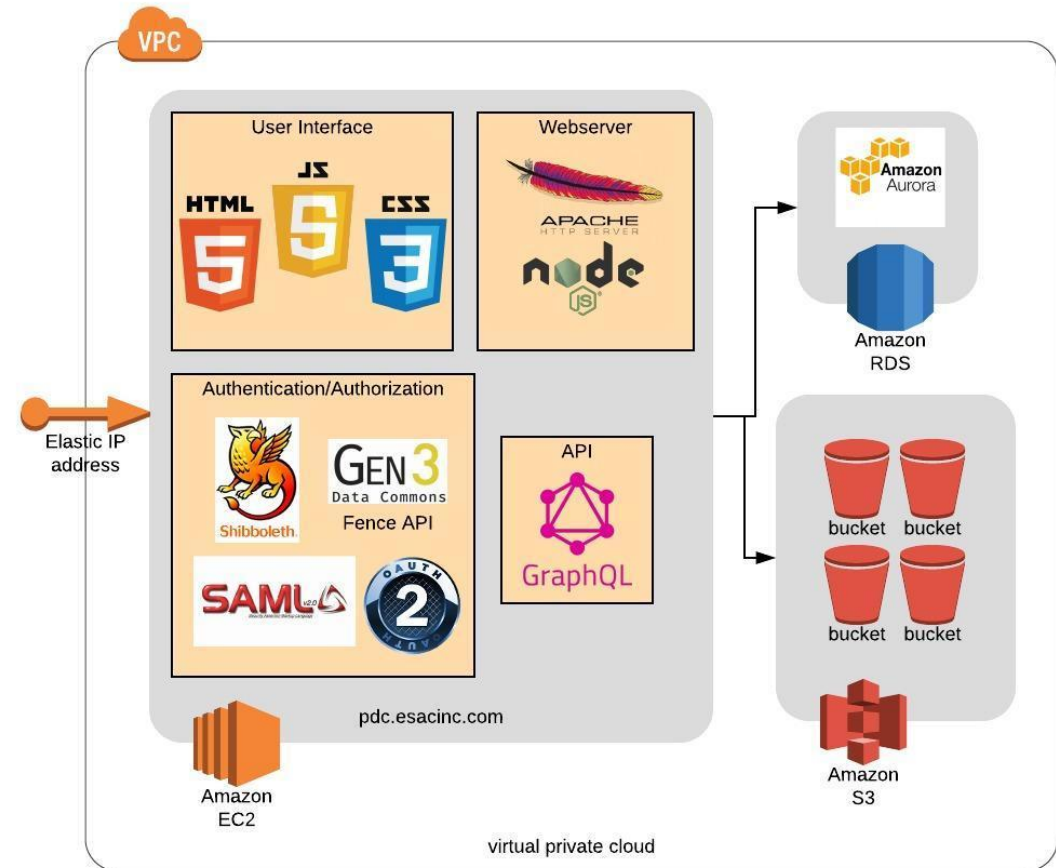


### PRIMARY SITES

15 cancer types from 11 primary sites

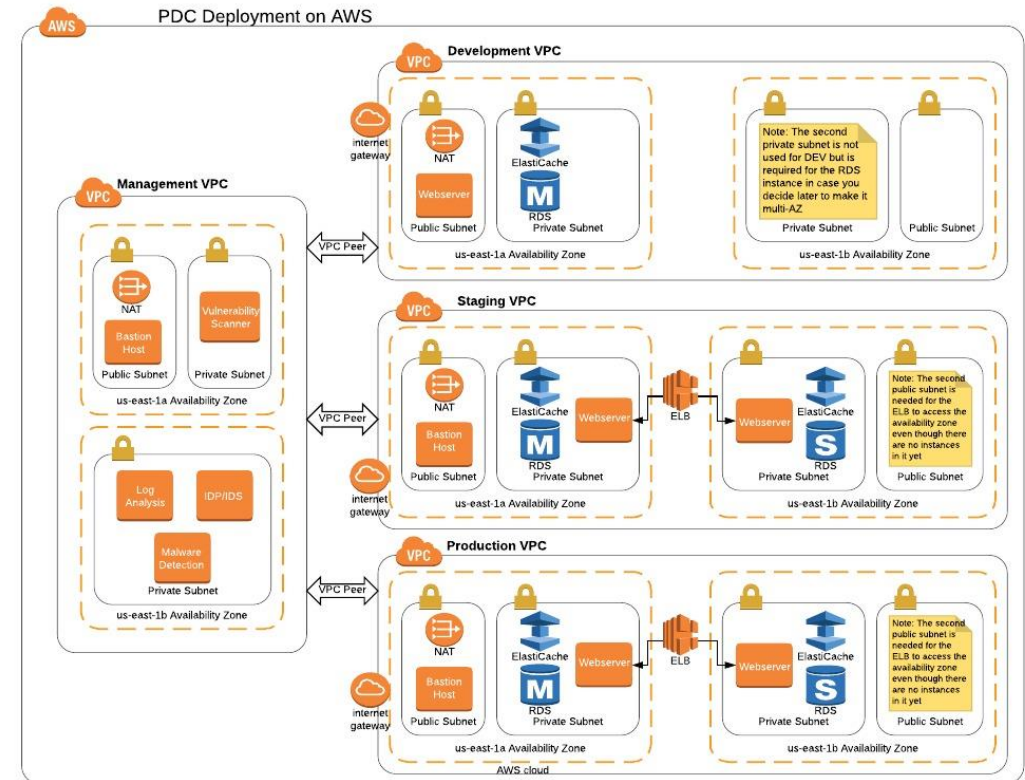
## PDC Technology Stack

- PDC Data Portal is a web-based application for querying and viewing Proteomic data
  - Also provides an API for programmatic access
- PDC Workspace is a web-based application for submitting raw data to the PDC for processing
  - Requires authentication (supports Google, NIH/eRA)
- Both work through any browser

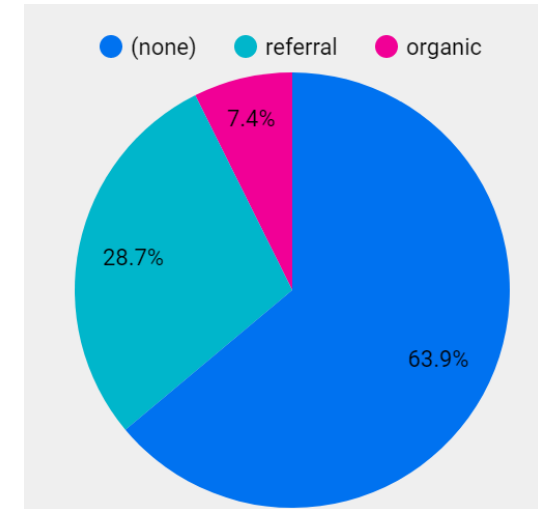
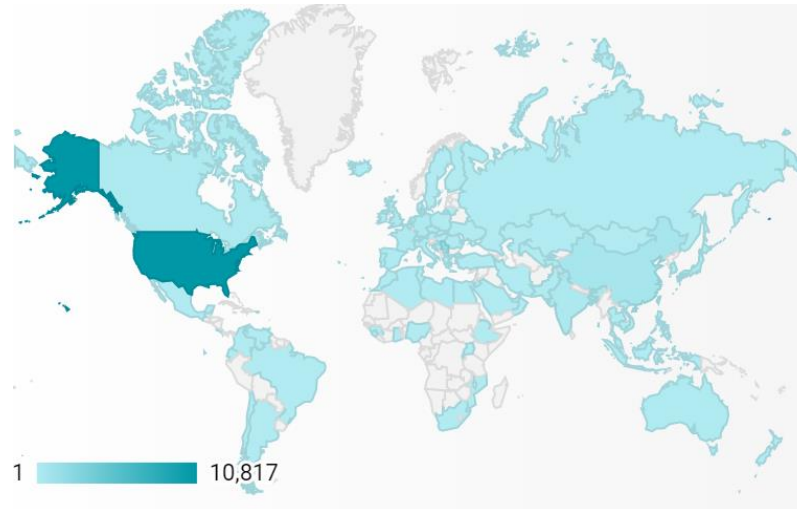
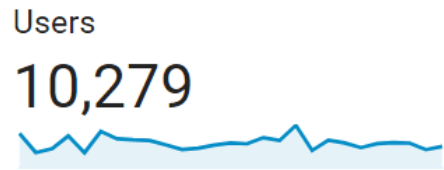


# PDC Infrastructure

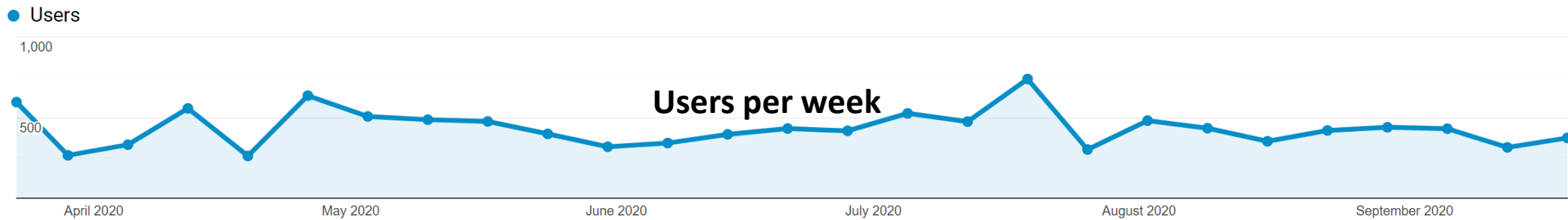
- PDC is FedRAMP and FISMA complaint system with Authority To Operate (ATO) under cancer.gov domain issued on January 15, 2020
  - 278 Security controls across 26 operational areas were planned and implemented
- PDC deployment is based on NIST architecture
- PDC went live on March 23, 2020



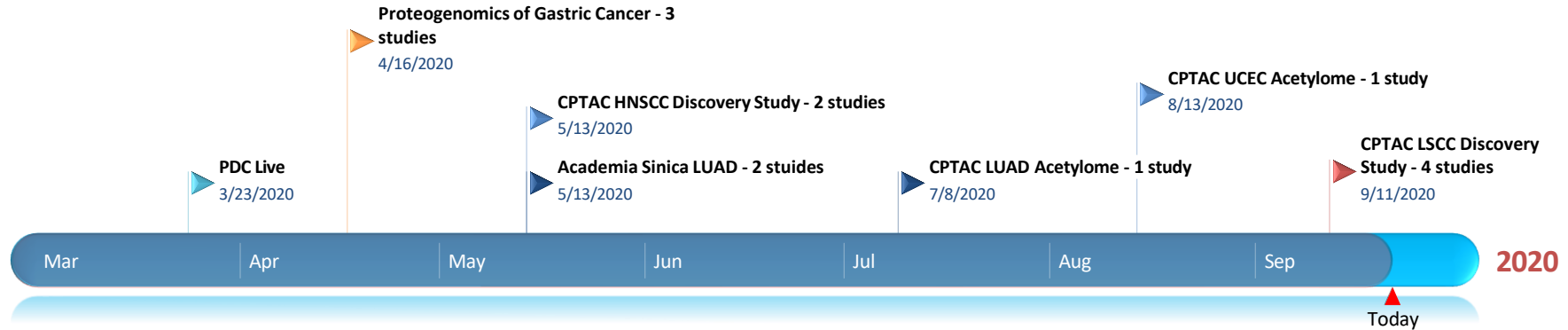
# PDC Analytics – since March 23, 2020



Traffic Source



## PDC timeline since launch



### Release includes

- raw data
- clinical data, experimental design
- processed data – open formats, PSM, Protein parsimony
- post processing for visualization
- API access

← → ↻ pdc.cancer.gov/pdc/browse

⌵ FILTERS

^ Primary Site

- Brain (4 Studies)
- Breast (4 Studies)
- Bronchus and lung (8 Studies)
- Colon (4 Studies)
- Head and Neck (2 Studies)
- Kidney (4 Studies)
- Liver (2 Studies)
- Lung (3 Studies)
- Not Reported (43 Studies)
- Ovary (7 Studies)
- Rectum (1 Study)
- Stomach (3 Studies)
- Uterus, NOS (3 Studies)

GENERAL

BIOSPECIMEN

CLINICAL

FILES

GENES

^ Program

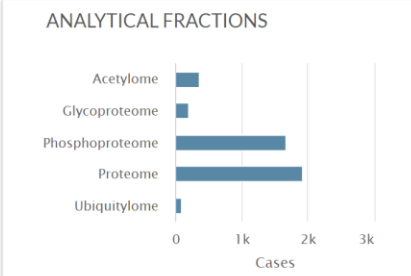
- Clinical Proteomic Tumor Analysis Consortium (42 Studies)
- Georgetown Proteomics Research Program (1 Study)
- International Cancer Proteogenome Consortium (7 Studies)
- Pediatric Brain Tumor Atlas - CBTT (2 Studies)
- Quantitative digital maps of tissue biopsies (1 Study)

^ Disease Type

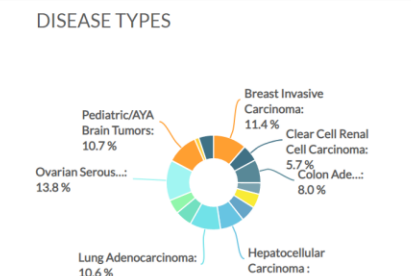
« Start searching by selecting a facet

^ Charts

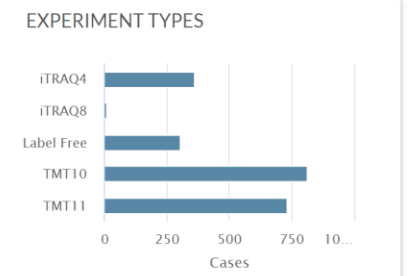
ANALYTICAL FRACTIONS



DISEASE TYPES



EXPERIMENT TYPES



Studies (53)   Biospecimens (3162)   Clinical (2045)   Files (76200)   Genes (14993)

Total studies: 53

Export All Manifests

Export Study Manifest

PDC Study ID	Study	Embargo Date	Project	Program	Disease Type	Primary Site	Analytical Fraction	Experiment Type	Cases #	Available files for data category					
										Raw	mzML	Metadat	PSM	Protein Assembly	Quality Metrics
<input type="checkbox"/> PDC000237	CPTAC LSCC Discovery Study - Ubiquitylome	2021-12-01	CPTAC3-Discovery	Clinical Proteomic Tumor Analysis Consortium	Other:Lung Squamous Cell Carcinoma	Not Reported;Bronchus and lung	Ubiquitylome	TMT11	89	30	30	11	60	6	2
<input type="checkbox"/> PDC000234	CPTAC LSCC Discovery Study - Proteome	2021-12-01	CPTAC3-Discovery	Clinical Proteomic Tumor Analysis Consortium	Lung Squamous Cell Carcinoma;O ther	Bronchus and lung;Not Reported	Proteome	TMT11	115	550	550	11	1100	5	2
<input type="checkbox"/> PDC000233	CPTAC LSCC Discovery Study - Acetyloyme	2021-12-01	CPTAC3-Discovery	Clinical Proteomic Tumor Analysis Consortium	Lung Squamous Cell Carcinoma;O ther	Bronchus and lung;Not Reported	Acetyloyme	TMT11	115	88	88	11	176	6	2
<input type="checkbox"/> PDC000232	CPTAC LSCC Discovery Study - Phosphoproteome	2021-12-01	CPTAC3-Discovery	Clinical Proteomic Tumor Analysis Consortium	Lung Squamous Cell Carcinoma	Bronchus and lung;Not Reported	Phosphoproteome	TMT11	115	286	286	11	572	6	2



## STUDY SUMMARY: CPTAC UCEC Discovery Study - Proteome

123

Cases

154

Aliquots

### SUMMARY

PDC Study Identifier	PDC000125	<b>Embargo Release Date</b>	<b>June 1, 2019</b>
Study ID	c935c587-0cd1-11e9-a064-0a9c39d33490	Analytical Fraction	Proteome
Study Name	CPTAC UCEC Discovery Study - Proteome	Disease Types	Uterine Corpus Endometrial Carcinoma
Experimental Strategy	TMT10	Project ID	CPTAC3-Discovery

Description ?

Protocol ?

Experimental Design ?

Clinical ?

Biospecimens ?

Workflow ?

DUA ?

### Data Use Agreement

CPTAC requests that data users abide by the same principles that were previously established in the Fort Lauderdale and Amsterdam meetings. The recommendations from the Fort Lauderdale meeting (2003) on best practices and principles for [sharing large-scale genomic data](#) address the roles and responsibilities of data producers, data users and funders of community resource projects. The aim of the recommendations is to establish and maintain an appropriate balance between the interests that data users have in rapid access to data and the needs that data producers have to publish and receive recognition for their work. The conclusion of the attendees at the Fort Lauderdale meeting was that a "responsible use" approach for secondary data users would be sufficient to ensure that the efforts of data producers will be recognized. "Responsible use" was defined as allowing data producers to have the opportunity to publish the initial global analyses of the data within a reasonable period of time prior to secondary analyses.

In 2008, the NCI's OCCPR organized a workshop to discuss how and when proteomics data should be released. The result was the [Amsterdam Principles](#) that established guidelines for the timing of data release, comprehensiveness of a dataset, data format, deposition to repositories, quality metrics, and responsibility for proteomic data release. Participants agreed that mass spectrometry output data files should be available to support the claims of proteomics publications. In 2010, NCI's OCCPR convened a follow-on workshop to address quality metrics for proteomics with an emphasis on mass spectrometry. As a sign of solidarity for these principles, four peer-reviewed journals simultaneously published the [corollary to Amsterdam Principles](#).

Agreeing to abide by these principles and the CPTAC Publication Guidelines is required to gain access to CPTAC data.

#### Common Data Analysis Pipeline (PDC Harmonization) data ?

Data Category	Files (n=1639)
Peptide Spectral Matches (Open Standard)	408
Peptide Spectral Matches (Text)	408
Processed Mass Spectra (Open Standard)	408
Protein Assembly (Text)	5
Quality Metrics (Text)	1
Quality Metrics (Web)	1
Raw Mass Spectra (Proprietary)	408

#### External References

Clinical Proteomic Tumor Analysis Consortium	S043
Database of Genotypes and Phenotypes	phs001287

#### Related PDC studies

#### Supplementary data ?

Data Category	Files (n=16)
Alternate Processing Pipeline (Archive)	3
Other Metadata (Document)	13



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

### PUBLICATIONS



Yongchao Dou, Emily A. Kawaler, Daniel Cui Zhou, Tao Liu, David Fenyo, et al., Cell (2020). Vol. 180, Issue 4, p729–748  
<https://doi.org/10.1016/j.cell.2020.01.026>

## Data Model, Standards, Semantics

### Robust Data Model

- Representing the data in a structured manner that allows to collect and distribute data and metadata effectively and efficiently

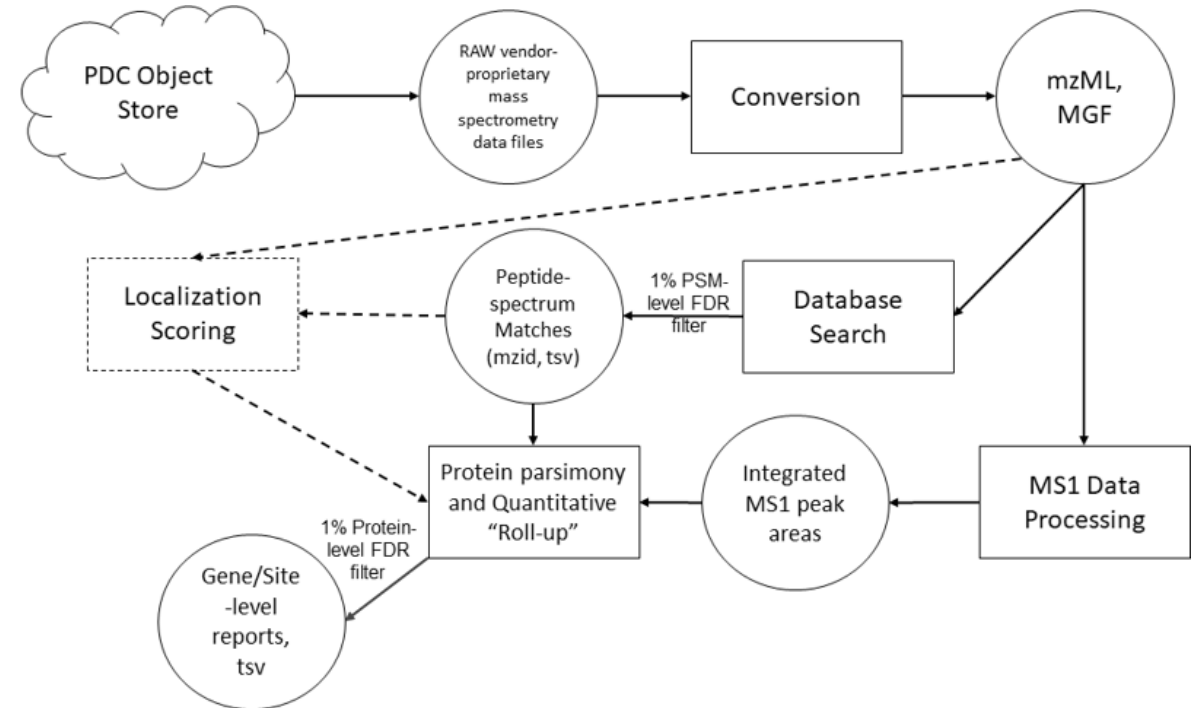
### Consistency of Metadata

- Reuse clinical metadata and standards from caDSR, NCI, ICD
- Use HUPO PSI Controlled Vocabulary of proteomics



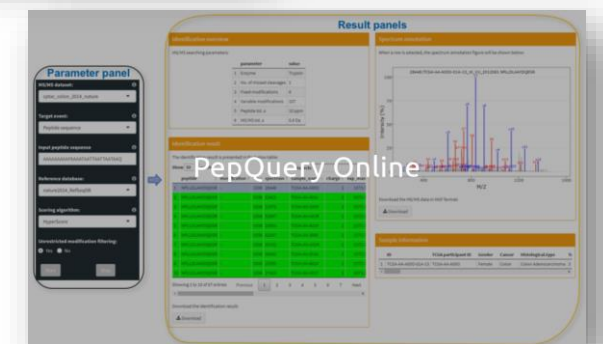
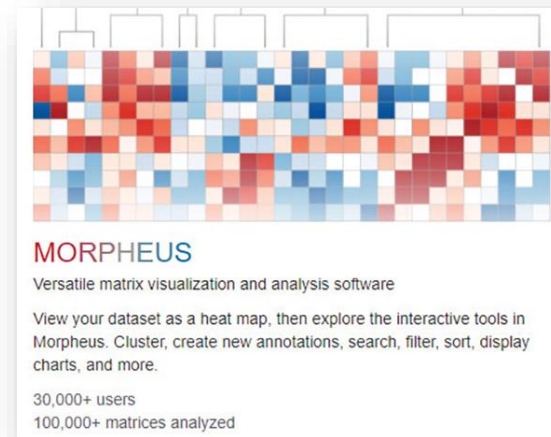
## Data harmonization

- Harmonization starts with assigning standard identifiers, data integrity checks, adherence to standards and PDC data model.
- All of the data is processed through a common data analysis pipeline (CDAP) for removing data analysis variables, enabling comparisons across datasets.



## PDC Analysis tools





- Morpheus viewer – a heatmaps visualization tool from Broad Institute for view expression data
- PepQuery - a peptide-centric search engine for novel peptide identification and validation from Bing Zhang's lab
- Jbrowse – a genome browser for viewing peptides on the genomic coordinates

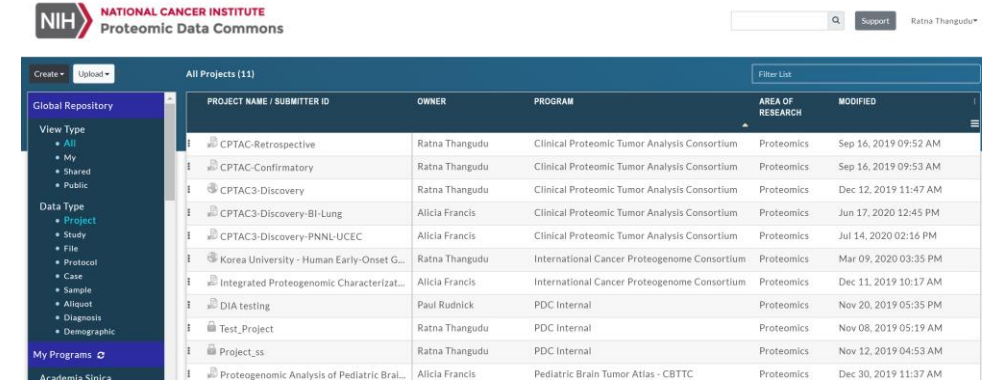


## PDC Data Submission

- A data submission template with examples is available
- A video tutorial and step by step guide
- UI menu driven and tsv files in predefined format
- Can request to set up a program or lab and have full control of the metadata
- Data remains private and modifiable until release

Join NCI Proteomic Data Portal

 Upload Files     Create Studies     Share Data     Analyze Data



NIH NATIONAL CANCER INSTITUTE Proteomic Data Commons





Create Upload All Projects (11) Filter List

PROJECT NAME / SUBMITTER ID	OWNER	PROGRAM	AREA OF RESEARCH	MODIFIED
CPTAC-Retrospective	Ratna Thangudu	Clinical Proteomic Tumor Analysis Consortium	Proteomics	Sep 16, 2019 09:52 AM
CPTAC-Confirmatory	Ratna Thangudu	Clinical Proteomic Tumor Analysis Consortium	Proteomics	Sep 16, 2019 09:53 AM
CPTAC3-Discovery	Ratna Thangudu	Clinical Proteomic Tumor Analysis Consortium	Proteomics	Dec 12, 2019 11:47 AM
CPTAC3-Discovery-BI-Lung	Alicia Francis	Clinical Proteomic Tumor Analysis Consortium	Proteomics	Jun 17, 2020 12:45 PM
CPTAC3-Discovery-PNNL-UCEC	Alicia Francis	Clinical Proteomic Tumor Analysis Consortium	Proteomics	Jul 14, 2020 02:16 PM
Korea University - Human Early-Onset G...	Ratna Thangudu	International Cancer Proteogenome Consortium	Proteomics	Mar 09, 2020 03:35 PM
Integrated Proteogenomic Characterizat...	Alicia Francis	International Cancer Proteogenome Consortium	Proteomics	Dec 11, 2019 10:17 AM
DIA testing	Paul Rudnick	PDC Internal	Proteomics	Nov 20, 2019 05:35 PM
Test_Project	Ratna Thangudu	PDC Internal	Proteomics	Nov 08, 2019 05:19 AM
Project_ss	Ratna Thangudu	PDC Internal	Proteomics	Nov 12, 2019 04:53 AM
Proteogenomic Analysis of Pediatric BraL...	Alicia Francis	Pediatric Brain Tumor Atlas - CBTT	Proteomics	Dec 30, 2019 11:37 AM

# Interoperability

- Data standards and harmonization – will help both general users and also the CCDH and CDA
- Large suite of GraphQL based APIs to provide flexibility for programmatic access
- Cross-referencing other multiomic resources
- Data files indexed by CRDC DCF service for easy access from analytical platforms such as SBG
- Authorization and authentication through DCF Fence API which allows users to register using their existing credentials such as Google, NHI, eRA, etc

*Cross references*

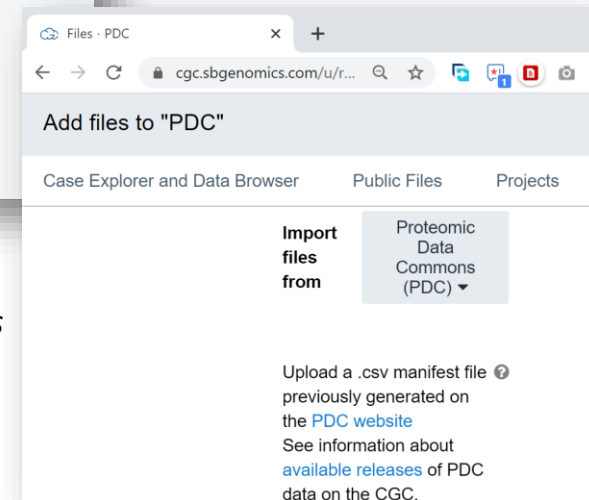
Case Submitter ID	Genomic and Imaging Data Resource
C3L-00413	 
C3I-00449	 

```

{
  geneSpectralCount(gene_name: "BRCC3") {
    gene_name
    NCBI_gene_id
    authority
    description
    organism
    chromosome
    locus
    proteins
    assays
    spectral_counts {
      study_submitter_id
      pdc_study_id
      gene_name
      distinct_peptide
      spectral_count
      unshared_peptide
      aliquot_id
    }
  }
}

```

*APIs*



Files · PDC

cgic.sbgenomics.com/u/r...

Add files to "PDC"

Case Explorer and Data Browser Public Files Projects

Import files from Proteomic Data Commons (PDC)

Upload a .csv manifest file previously generated on the [PDC website](#). See information about [available releases](#) of PDC data on the CGC.

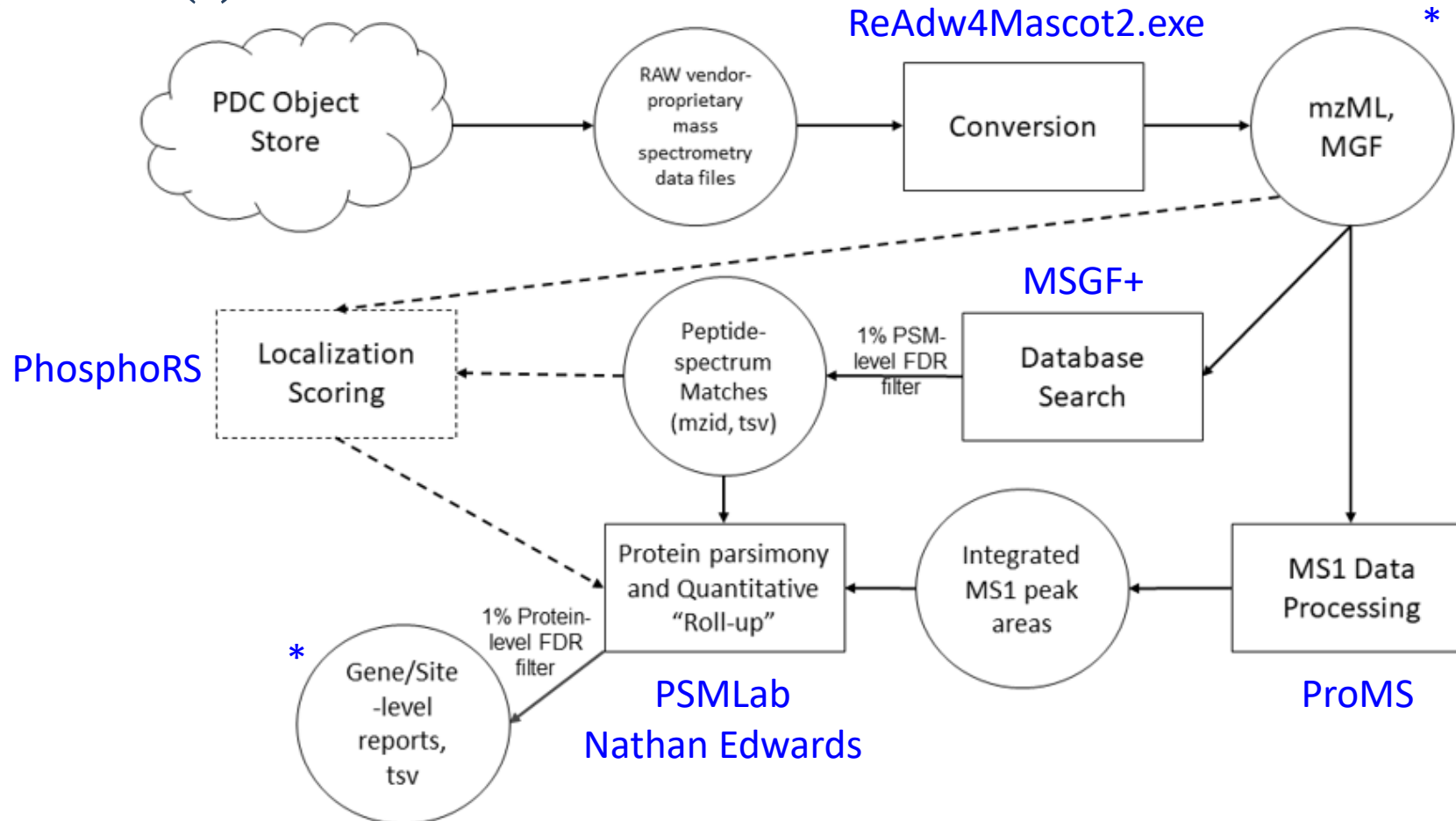
*Manifest uploads*

# Data harmonization, APIs, CDAP and other pipelines

---

Paul Rudnick

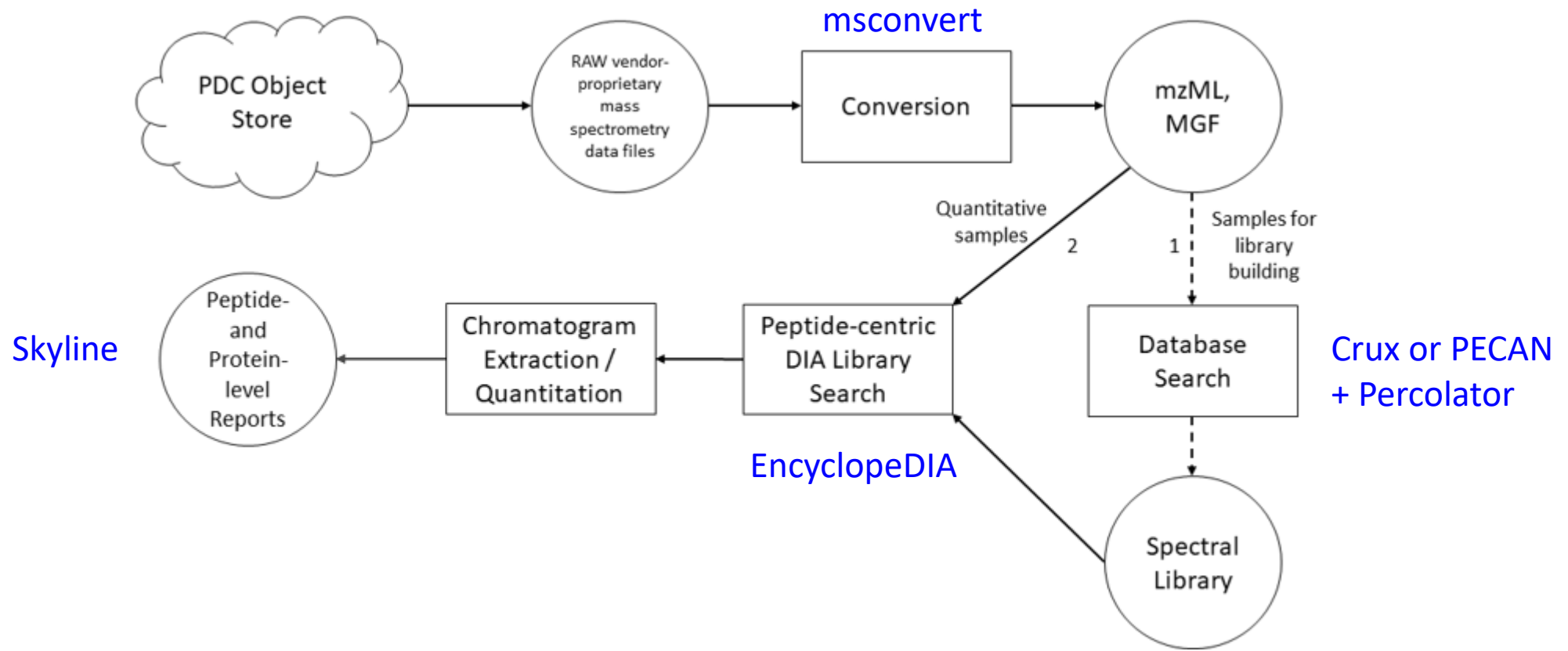
# Harmonization (DDA): Common Data Analysis Pipeline(s): 2013-current



\* Distributed



# Harmonization (DIA): Common Data Analysis Pipeline(s): 2018-current



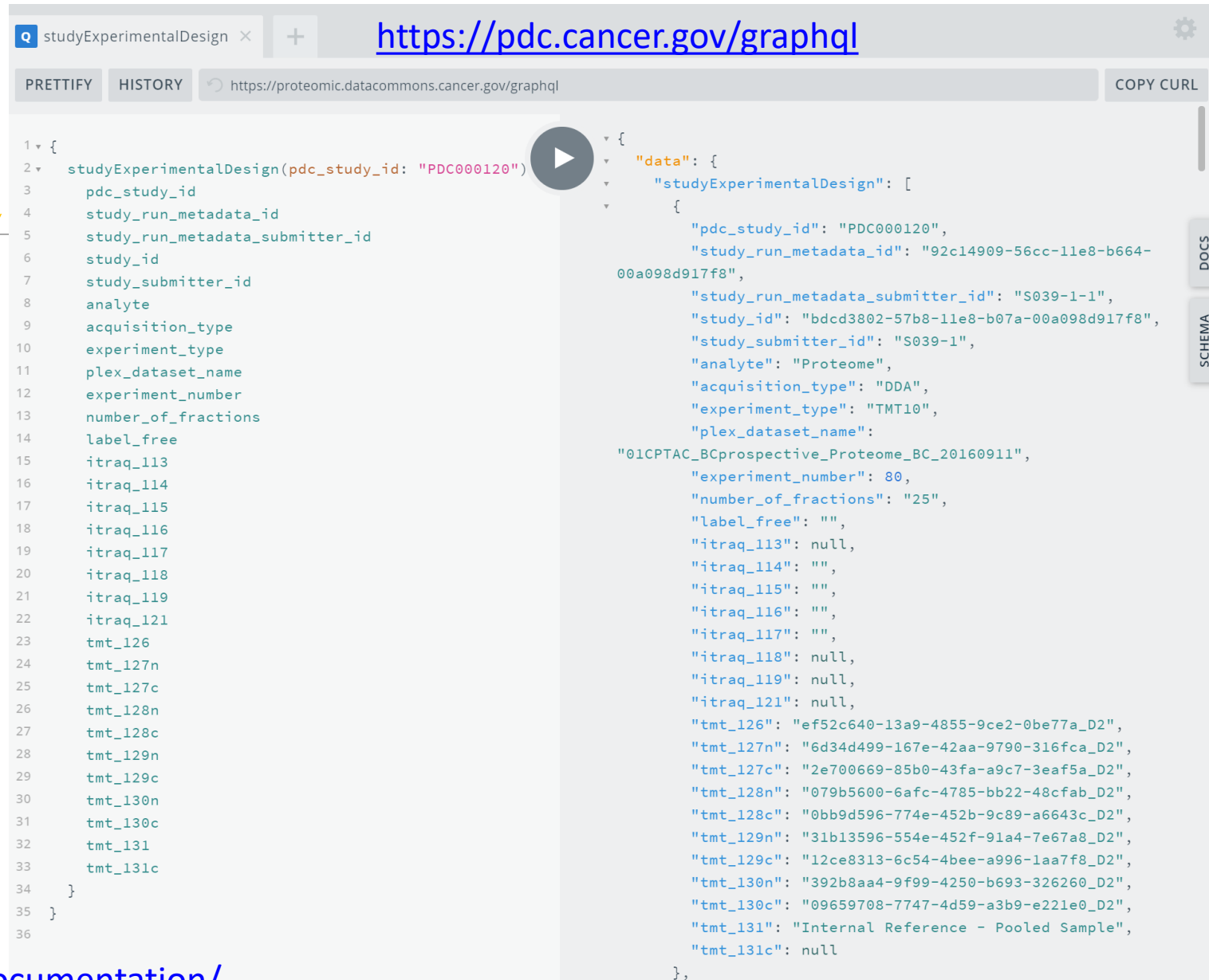
# How to aliquots map to TMT channels for a given study?

Gets experimental design for a PDC Study ID

Show samples </>

**GET** ?query={ studyExperimentalDesign (pdc\_study\_id: "{pdc\_study\_id}")  
{ pdc\_study\_id, study\_run\_metadata\_id,  
study\_run\_metadata\_submitter\_id, study\_id, study\_submitter\_id,  
analyte, acquisition\_type, experiment\_type, plex\_dataset\_name,  
experiment\_number, number\_of\_fractions, label\_free, itraq\_113,  
itraq\_114, itraq\_115, itraq\_116, itraq\_117, itraq\_118, itraq\_119,  
itraq\_121, tmt\_126, tmt\_127n, tmt\_127c, tmt\_128n, tmt\_128c,  
tmt\_129n, tmt\_129c, tmt\_130n, tmt\_130c, tmt\_131, tmt\_131c } }

Example from Swagger page.



The screenshot shows a browser window with the URL <https://pdc.cancer.gov/graphql>. The browser's address bar also shows <https://proteomic.datacommons.cancer.gov/graphql>. The page has tabs for 'PRETTIFY', 'HISTORY', and 'COPY CURL'. The main content area displays a GraphQL query on the left and its JSON response on the right. A play button icon is visible between the query and response.

```

1 {
2   studyExperimentalDesign(pdc_study_id: "PDC000120")
3     pdc_study_id
4     study_run_metadata_id
5     study_run_metadata_submitter_id
6     study_id
7     study_submitter_id
8     analyte
9     acquisition_type
10    experiment_type
11    plex_dataset_name
12    experiment_number
13    number_of_fractions
14    label_free
15    itraq_113
16    itraq_114
17    itraq_115
18    itraq_116
19    itraq_117
20    itraq_118
21    itraq_119
22    itraq_121
23    tmt_126
24    tmt_127n
25    tmt_127c
26    tmt_128n
27    tmt_128c
28    tmt_129n
29    tmt_129c
30    tmt_130n
31    tmt_130c
32    tmt_131
33    tmt_131c
34  }
35 }
36

```

```

{
  "data": {
    "studyExperimentalDesign": [
      {
        "pdc_study_id": "PDC000120",
        "study_run_metadata_id": "92c14909-56cc-11e8-b664-00a098d917f8",
        "study_run_metadata_submitter_id": "S039-1-1",
        "study_id": "bdcd3802-57b8-11e8-b07a-00a098d917f8",
        "study_submitter_id": "S039-1",
        "analyte": "Proteome",
        "acquisition_type": "DDA",
        "experiment_type": "TMT10",
        "plex_dataset_name": "01CPTAC_BCprospective_Proteome_BC_20160911",
        "experiment_number": 80,
        "number_of_fractions": "25",
        "label_free": "",
        "itraq_113": null,
        "itraq_114": "",
        "itraq_115": "",
        "itraq_116": "",
        "itraq_117": "",
        "itraq_118": null,
        "itraq_119": null,
        "itraq_121": null,
        "tmt_126": "ef52c640-13a9-4855-9ce2-0be77a_D2",
        "tmt_127n": "6d34d499-167e-42aa-9790-316fca_D2",
        "tmt_127c": "2e700669-85b0-43fa-a9c7-3eaf5a_D2",
        "tmt_128n": "079b5600-6afc-4785-bb22-48cfab_D2",
        "tmt_128c": "0bb9d596-774e-452b-9c89-a6643c_D2",
        "tmt_129n": "31b13596-554e-452f-91a4-7e67a8_D2",
        "tmt_129c": "12ce8313-6c54-4bee-a996-1aa7f8_D2",
        "tmt_130n": "392b8aa4-9f99-4250-b693-326260_D2",
        "tmt_130c": "09659708-7747-4d59-a3b9-e221e0_D2",
        "tmt_131": "Internal Reference - Pooled Sample",
        "tmt_131c": null
      }
    ]
  }
}

```

# Jupyter Notebook Example

## NATIONAL CANCER INS Proteomic Data

This notebook attempts to demonstrate the following:

- Using the Proteome Data Commons (PDC) API to retrieve relative protein express Common Data Analysis Pipeline (CDAP). More information on the PDC implementa
- Using the PDC API to retrieve the associated clinical metadata.
- Formatting the data for analysis.
- Clustering the data using the Seaborn clustermap package.
- Visualizing the clustermap / heatmap.

The results are intended to help identify clusters of samples (tumors) displaying similar

These are the required imports. Install them using pip if needed.

```
In [1]: import requests
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Next, set up the query parameters.

The first one is `study_submitter_id`. These can be retrieved using an API like this [one](#)

```
In [2]: study_submitter_id = 'S015-1' # S015-1 is TCGA-Breast(iTRAQ4)
```

Next, select the data type to retrieve for the given study. A table of data\_types is availa [here](#). In brief, these values are log2 transformed ratio of the sample to the control chan normalization.

```
In [3]: data_type = 'log2_ratio' # Retrieves CDAP iTRAQ or TMT data
```

Next, set the number of samples to retrieve. Samples are identified by their aliquot\_sut currently recommended during the initial PDC development period. Higher values may

```
In [4]: max_aliquots = 25
```

Next, the expression data GraphQL query is set up. Adding the `study_submitter_id` a

```
In [5]: exp_data_query = '''
{
  paginatedDataMatrix(study_submitter_id: ''' + study_submitter_id
    offset: 0 limit: ''' + str(max_aliquots) + ''' ) {
    total
    dataMatrix
    pagination {
      count
      sort
      from
      page
      total
      pages
      size
    }
  }
},...'''
```

## Get the TMT ra

Let's do the same thing for the clinical data.

```
In [6]: metadata_query = '''
{
  clinicalMetadata(study_submitter_id: ''' + study_submitter_id
    aliquot_submitter_id
    morphology
    primary_diagnosis
    tumor_grade
    tumor_stage
  }
},...'''
```

## Get the clinical

Now we can define a function to make the GraphQL Post query. This will get called one new to GraphQL, you can also try your queries [here](#).

```
In [7]: def query_pdc(query):
URL = 'https://pdc-dev.esacinc.com/graphql'
# Send the POST graphql query
print('Sending query.')
pdc_response = requests.post(URL, json={'query': query})

# Set up a data structure for the query result
decoded = dict()

# Check the results
if pdc_response.ok:
# Decode the response
decoded = pdc_response.json()
else:
# Response not OK, see error
pdc_response.raise_for_status()
return decoded
```

Retrieve the expression data and convert it into a pandas dataframe.

```
In [8]: decoded = query_pdc(exp_data_query)
matrix = decoded['data']['paginatedDataMatrix']['dataMatrix']

# Aliquots are first row, gene names are first column
ga = pd.DataFrame(matrix[1:], columns=matrix[0]).set_index('Gene/Aliquot')
print('Created a dataframe of these dimensions: {}'.format(ga.shape))
```

Sending query.  
Created a dataframe of these dimensions: (10625, 25)

Since the expression values are returned as strings, we need to convert those to floats and deal with missing data.

```
In [9]: for col in ga.keys():
ga[col] = pd.to_numeric(ga[col], errors='coerce')
```

The clustermap module within the Seaborn package does not allow for NaN values. So we must create a mask value that does not interfere much with the clustering and is likely to be unique. Not imputation is used. Missing data is a particularly tough challenge for proteomics data, particularly for phosphorylation studies. By using a value close to 0, we are saying that these are unchanged between samples. Better solutions may be used.

```
In [11]: decoded = query_pdc(metadata_query)
matrix = decoded['data']['clinicalMetadata']
metadata = pd.DataFrame(matrix, columns=matrix[0]).set_index
print('Created a dataframe of these dimensions: {}'.format(m
```

Sending query.  
Created a dataframe of these dimensions: (111, 4)

We can then set up a color mapping function for the clinical annotations.

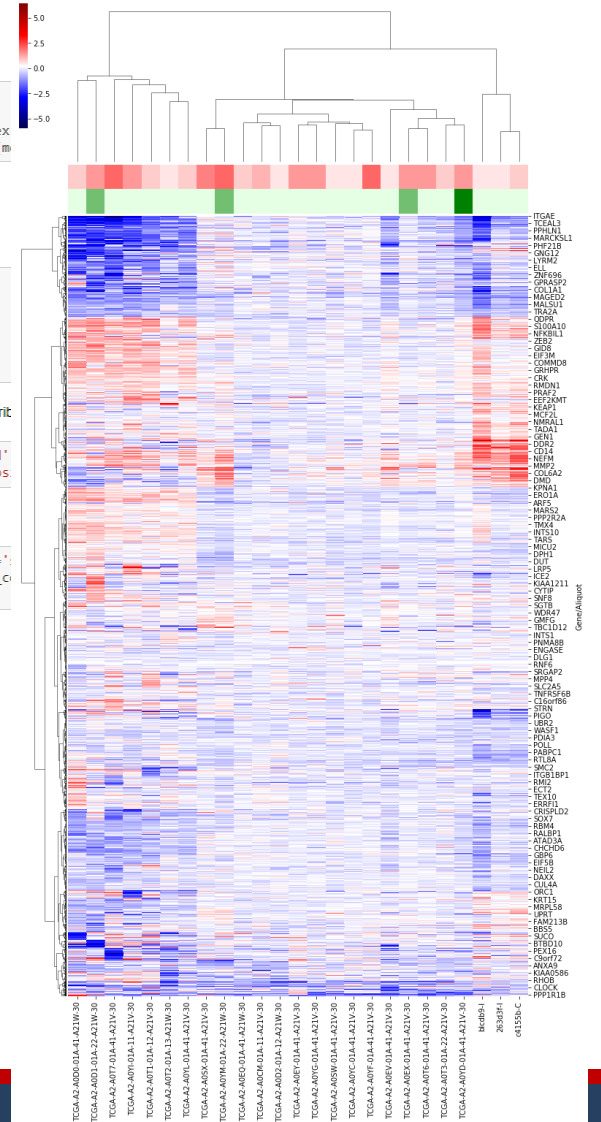
```
In [12]: def get_colors(df, name, color) -> pd.Series:
s = df[name]
su = s.unique()
colors = sns.light_palette(color, len(su))
lut = dict(zip(su, colors))
return s.map(lut)
```

Next, call `get_colors()` to map the `tumor_stage` and `primary_diagnosis` attrit

```
In [13]: stage_col_colors = get_colors(metadata, 'tumor_stage', 'red')
diagnosis_col_colors = get_colors(metadata, 'primary_diagnos
```

And, finally, generate the large clustermap.

```
In [14]: sns.clustermap(ga, metric='euclidean', method='ward', cmap='
col_colors=[stage_col_colors, diagnosis_col_col
plt.show()
```



# Seven Bridges Integration

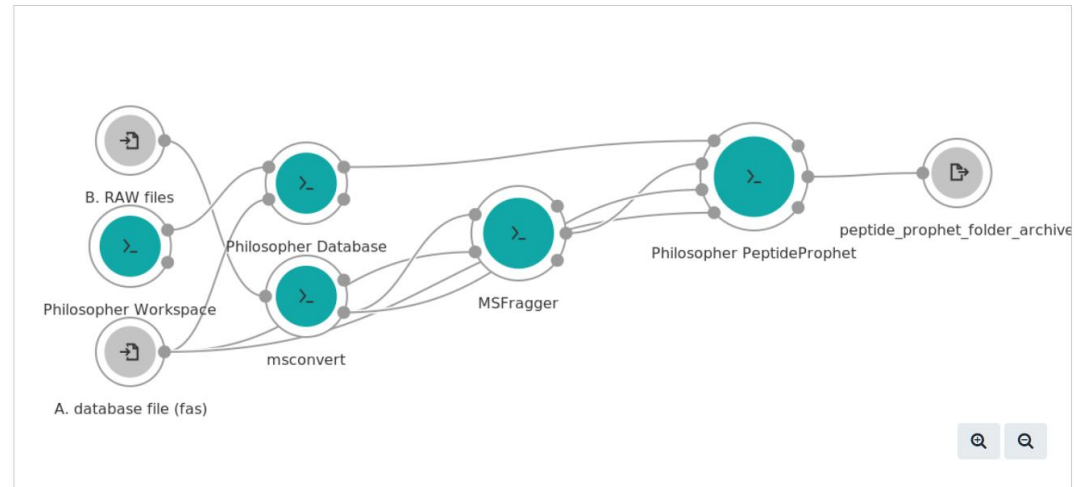
- Direct loading of files by manifest generated on the PDC (i.e., no upload/download of data files)
- Applications added as Docker containers
- Pipelines designed using Rabbix (GUI for generating CWL)
- MSFragger + Prophets + TMT analysis of ccRCC dataset 23 plexes -> \$13.31

## Philosopher - Convert to Peptide Prophet

Created by [david.roberson](#) on May 4, 2020 11:44 • Last edited by [david.roberson](#) on June 7, 2020 11:55

### Description

No description.



### Ports

ID	Label	Type	Required	Prefix	Format
database_name	A. database file (fas)	File	Yes		FAS
raw_files	B. RAW files	File array	Yes		RAW

Revision 114 [Edit](#) [Run](#) [...](#)

### Basic Information

CWL Version [?](#) v1.0

Contributors: [prvst](#) [david.roberson](#)

App Id: david.roberson/philosopher-dev/msconvert-wf

### Workflow steps >

[msconvert >](#)


[MSFragger >](#)

[Philosopher Workspace >](#)


[Philosopher Database >](#)

[Philosopher PeptideProphet >](#)


**Members** [Email notifications](#)




[david.roberson](#) OWNER  
Write, Copy, Execute, Admin



[paul.rudnick](#) ADMIN  
Write, Copy, Execute, Admin



[prvst](#) ADMIN  
Write, Copy, Execute, Admin



[kghose](#)  
Write, Copy, Execute

[Manage members](#)
[Leave project](#)

Dave Roberson  
Manisha Ray  
Felipe Lprevost

# PDC Quant Data Loaded into Google Big Query Tables @ ISB-CGC

ISB-CGC Data Browsers Resources Docu

Get started today!

## BigQuery Table Search

ISB-CGC BigQuery Documentation ISB-CGC BigQuery Access Info Google BigQuery Console About BigQuery

Explore and learn more about available ISB-CGC BigQuery tables with this search feature. Find tables of interest based on category, reference genome build, data type and free-form text search.

**Data Browsers**

### BigQuery Table Search

Browse BigQuery table metadata and molecular cancer data from the Genomic Data Commons and other sources. Jump directly to a table to perform discovery and computation via SQL.

Learn Launch

**Status**  
CURRENT

**Name**

**Program**  
Choose Programs...

**Category**

- CLINICAL BIOSPECIMEN DATA
- FILE METADATA
- GENOMIC REFERENCE DATABASE
- PROCESSED -OMICS DATA

**Reference Genome**  
ALL

**Source**  
PDC

Show 10 entries

Columns CSV Download Search:

Name	Program	Category	Source	Data Type	Status	Rows	Created	Preview	Open
TCGA-BRCA ITRAQ PHOSPHOPEPTIDE QUANTITATION REPORT	CPTAC	PROCESSED -OMICS DATA	PDC	PROTEIN EXPRESSION	CURRENT	2,511,668	5/17/2017		
TCGA-BRCA ITRAQ PHOSPHOSITE QUANTITATION REPORT	CPTAC	PROCESSED -OMICS DATA	PDC	PROTEIN EXPRESSION	CURRENT	2,345,559	5/17/2017		
TCGA-BRCA ITRAQ PROTEOME QUANTITATION REPORT	CPTAC	PROCESSED -OMICS DATA	PDC	PROTEIN EXPRESSION	CURRENT	1,007,311	5/18/2017		
TCGA-BRCA PUBMED-ID 27251275 - SUPPTABLE01	CPTAC	PROCESSED -OMICS DATA	PDC	PROTEIN EXPRESSION	CURRENT	105	3/21/2017		
TCGA-OV ITRAQ JHU PROTEOME QUANTITATION REPORT	CPTAC	PROCESSED -OMICS DATA	PDC	PROTEIN EXPRESSION	CURRENT	947,385	5/18/2017		
TCGA-OV ITRAQ PNNL PROTEOME QUANTITATION REPORT	CPTAC	PROCESSED -OMICS DATA	PDC	PROTEIN EXPRESSION	CURRENT	500,239	5/18/2017		

Showing 1 to 6 of 6 entries (filtered from 528 total entries)

Previous 1 Next

Have feedback or corrections? Please email us at [feedback@isb-cgc.org](mailto:feedback@isb-cgc.org).

Kawther Abdilleh  
Bill Longabaugh

Our Contact Information:

- Mike MacCoss, UW ([maccoss@uw.edu](mailto:maccoss@uw.edu))
- R Rajesh Thangudu, ESAC Inc. ([ratna.thangudu@esacinc.com](mailto:ratna.thangudu@esacinc.com))
- [nci.pdc.help@esacinc.com](mailto:nci.pdc.help@esacinc.com)