

## I. Biospecimens

### TCGA Ovarian Cancer Tumor Samples

TCGA ovarian cancer tumor samples have a human readable bar code and a Universally Unique Identifier (UUID). Both identifiers are tracked in the TCGA Data Portal Biospecimen Metadata Browser (<https://tcga-data.nci.nih.gov/uuid/uuidBrowser.htm>) and in the CPTAC Data Portal.

#### Human Readable Bar Code

The human readable bar code is composed of 7 parts (codes) separated by a hyphen (-)

Example: TCGA-24-1104-01A-31-2224-29

1. TCGA Project
2. 24 Tissue source site, location where the samples and clinical metadata were collected
3. 1104 Study participant
4. 01A Sample type and vial, 01 = a solid tumor, A = The first vial in a sequence of samples
5. 31 Portion, order of portion in a sequence of 100 – 120 mg samples, here it is portion 11
6. 2224 Plate, order of plate in a sequence of 96-well plates
7. 29 Center that will receive the sample, Center 29 = Pacific Northwest National Lab (PNNL)

Two CPTAC Proteome Characterization Centers (PCCs) worked on the TCGA Ovarian Cancer Proteome Study, some of the bar codes will have the Center number 28.

Example: TCGA-24-2024-01A-41-2222-28

Center 28 = The Johns Hopkins University Proteomics

The codes (parts) in the human readable identifier can be looked up in the Code Table Report <https://tcga-data.nci.nih.gov/datareports/codeTablesReport.htm?codeTable=center>

More information about the TCGA bar codes can be found on the National Cancer Institute wiki page, <https://wiki.nci.nih.gov/display/TCGA/TCGA+Barcode>

## Universally Unique Identifier (UUID)

The UUID is an identifier standard used in software construction, standardized by the Open Software Foundation (OSF) as part of the Distributed Computing Environment (DCE). A UUID is a 16-byte (128-bit) hexadecimal value. In its canonical form, a UUID consists of 32 hexadecimal digits, displayed in 5 groups separated by hyphens, in the form 8-4-4-4-12 for a total of 36 characters (32 digits and 4 hyphens). For example:

67617e82-a476-4f70-8aa8-df3cccc7dcdd

Information on TCGA UUIDs can be found on this page,  
<https://wiki.nci.nih.gov/display/TCGA/UUID+Migration+Plan>

The TCGA bar codes and UUIDs are listed in the file, “OV\_All\_clinical\_features\_TCGAAbiotab\_CPTAC\_S020.” Here you will see the TCGA bar code and UUID of the sample that was shipped to the PCC and also the patient TCGA bar code and UUID.

The TCGA bar code for the sample is used in the file, “CPTAC\_TCGA\_Ovarian\_Cancer\_iTRAQ\_Sample\_Mapping.”

## TCGA Ovarian Cancer Mixed Tumor Samples

Internal reference samples for iTRAQ experiments were prepared at each PCC.

PNNL Pooled sample of 84 TCGA ovarian tumor tissue samples, 117C  
JHU Pooled sample of 72 TCGA ovarian tumor tissue samples, 114C

## Ovarian Tumor Tissue Sample , JHU iTRAQ Datasets

JHU included an ovarian tumor tissue sample (not from the TCGA program) in the proteome and glycoproteome analyses. This biospecimen is labeled as OVARIAN-CONTROL, noted in both Folder and File names.

## **II. Folder Names**

Folder or Data Set names include the 3 TCGA samples assayed in the iTRAQ experiment. Seven elements are present, separated by an underscore ( \_ ).

Example:

TCGA\_25-1316-01A\_24-1550-01A\_24-1555-01A\_Proteome\_PNNL\_20130303

1. TCGA Project

2. 25-1316-01A Sample used in the iTRAQ experiment only part of the bar code, tissue source site (25), study participant (1316), and solid tumor and vial (01A)
3. 24-1550-01A Sample used in the iTRAQ experiment
4. 24-1555-01A Sample used in the iTRAQ experiment
5. Proteome Type of analysis (other options are Glycoproteome, Phosphoproteome)
6. PNNL Site generating the mass spectrometry file
7. 20130208 Date yyyyymmdd format

Within the Data Set folder are subfolders for each of the data types

(**raw**) the original mass spectrometry(MS) instrument files;

(**mzML**) HUPO-PSI standard raw data files generated from the original MS instrument files;

(**PSM**): Peptide-Spectrum Match data (includes .tsv and .mzIdentML folders)

### III. File Names

File names contain descriptive parameters separated by an underscore ( \_ ), fields include TCGA biospecimen partial names and laboratory-specific information designated by the PCCs.

#### PNNL File Name Examples

Proteome

TCGA\_25-1316\_24-1550\_24-1555\_117C\_W\_PNNL\_B1S1\_f01.raw

Phosphoproteome

TCGA\_25-1316\_24-1550\_24-1555\_117C\_P\_PNNL\_B1S1\_f01.raw

1. TCGA Project
2. 25-1316 Sample used with the isobaric reagent mass = 114, only tissue source site (25) and study participant (1316) are included
3. 24-1550 Sample used with the isobaric reagent mass = 115
4. 24-1555 Sample used with the isobaric reagent mass = 116
5. 117C Sample used with the isobaric reagent mass = 117, Pooled sample of 84 TCGA ovarian tumor tissue samples, internal reference
6. W, P (W) whole, global proteome, (P) phosphoproteome
7. PNNL Mass Spectrometry Site
8. B1S1 PCC- specific reference
9. f01 Fraction number

### JHU File Names Examples

#### Proteome

TCGA\_114C\_13-1492-01A-01\_36-2530-01A-01\_25-2409-01A-01\_W\_JHUZ\_20130228\_F01.raw

#### Glycoproteome

TCGA\_114C\_13-1492-01A-01\_36-2530-01A-01\_25-2409-01A-01\_G\_JHUZ\_20130228\_RUN1\_NOFRACTION.raw

1. TCGA Project
2. 114C Sample used with the isobaric reagent mass = 114, Pooled sample of 72 TCGA ovarian tumor tissue samples
3. 13-1492-01A-01 Sample used with the isobaric reagent mass = 115 only partial TCGA bar code, includes tissue source site (13), study participant (1492), Sample type and vial (01A) \*\*\* final parameter (01) PCC specific reference number, not tumor sample portion number\*\*\*
4. 13-1492-01A-01 Sample used with the isobaric reagent mass = 116
5. 25-2409-01A-01 Sample used with the isobaric reagent mass = 117
6. W, G (W) whole, global proteome, (G) glycoproteome
7. JHUZ Mass Spectrometry Site
8. 20130224 Date yyymmdd format
9. F01 Fraction number

### **IV. Data Types Available for Download on the Study Page**

(ALL): Selection of this box downloads all data in the row

(raw): The original mass spectrometry(MS) instrument files

(mzML): HUPO-PSI standard raw data files generated from the original MS instrument files

(PSM): Peptide-Spectrum Match data from the Common Data Analysis Pipeline (CDAP)

(meta): Clinical data files, mapping of biospecimens to iTRAQ labels (where applicable), folder and file naming conventions

Checksum files are included in all downloads for verification

### **V. Resources**

CPTAC Public Data Portal: "About the Data" tab, "Data Levels and Data Types" section