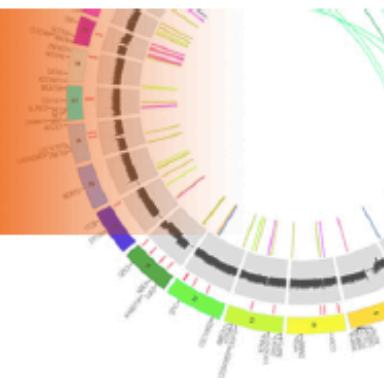


# TCGA: An Overview

*February 14<sup>th</sup>, 2013*

*Jay Bowen, MS, PMP  
 Program Director  
 Biospecimen Core Resource*

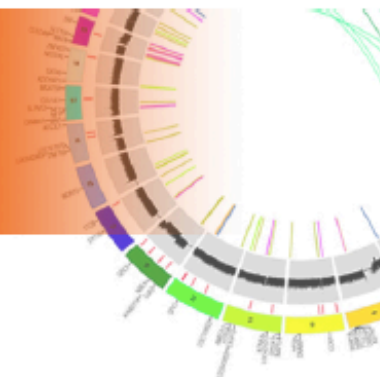
# The Cancer Genome Atlas (TCGA)



**The Cancer Genome Atlas (TCGA)** is a project of the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI) to develop a comprehensive molecular atlas of cancer.

TCGA's integrated database of molecular and clinical annotation will provide scientists with accelerated access to the discovery with the end goal being better diagnostics, therapies, and preventive measures for cancer.

# TCGA General Eligibility Criteria



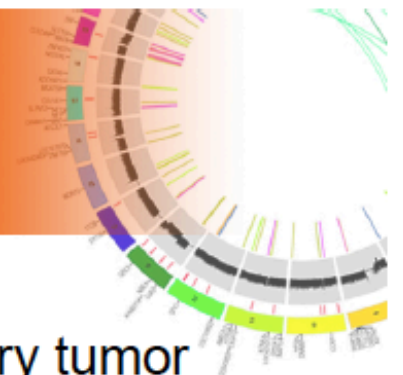
- **General Biospecimen Criteria**

- Primary, *de novo* tumor
- Invasive cancer, not *in situ*
- Metastatic and/or recurrences encouraged in addition to primary tumor
- Diagnosis consistent with TCGA list of eligible tumors and subtypes
- Prior malignancies may not have had prior systemic chemotherapy or local radiation in same field
- Tumor Nuclei  $\geq 60\%$ ; Tumor necrosis  $\leq 20\%$
- Blood as germline control for most cases

- **Additional Requirements**

- Complete, de-identified pathology report in English
- Completion of required clinical data elements
- Submission of pre-screen frozen sections and diagnostic slides

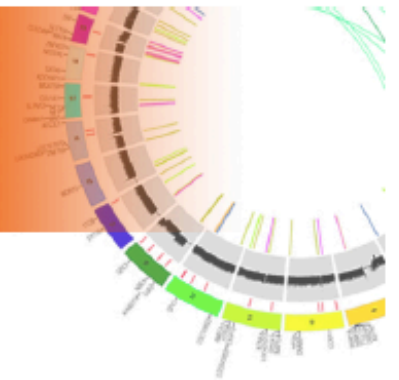
# Eligibility Exceptions



- Metastatic melanoma tumors accepted if no primary tumor
- ‘Acceptable’ prior treatments
  - BCG for bladder
  - Interferon for melanoma
  - Hydroxyurea for acute myeloid leukemia



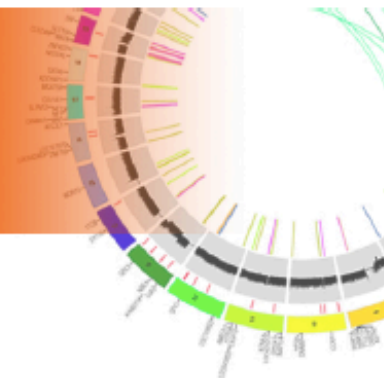
# TCGA Tumor Studies – Open



- Acute Myeloid Leukemia
- Bladder Urothelial Carcinoma
- Breast Invasive Carcinoma\*
- Cervical Carcinoma
- Diffuse Large B-Cell Lymphoma
- Esophageal Carcinoma
- Head and Neck Carcinoma
- Hepatocellular Carcinoma
- Kidney Papillary Carcinoma
- Lower Grade Glioma
- Melanoma
- Pancreatic Adenocarcinoma
- Prostate Adenocarcinoma
- Stomach Adenocarcinoma
- Sarcoma

\* *Open only to non-ductal histologies*

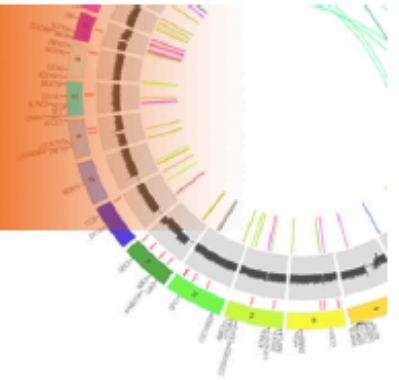
# TCGA Tumor Studies – Closed



- Colon Adenocarcinoma
- Endometrial Carcinoma
- Glioblastoma Multiforme
- Kidney Clear Cell Carcinoma
- Lung Adenocarcinoma
- Lung Squamous Cell Carcinoma
- Ovarian Serous Cystadenocarcinoma
- Rectal Adenocarcinoma
- Thyroid Papillary Carcinoma

NOTE - All studies are open to cases from patients of African descent

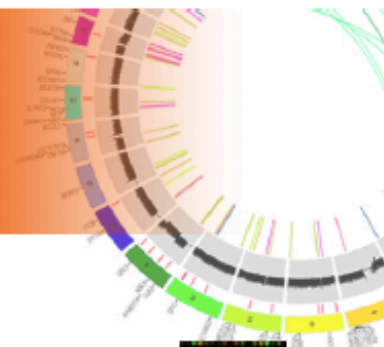
# TCGA Rare Tumor Studies



- Adrenocortical Carcinoma
- Chromophobe Kidney\*
- Mesothelioma
- Pheochromocytoma/Paraganglioma
- Uterine Carcinoma\*
- Acute Lymphoblastic Leukemia
- Anaplastic Thyroid
- Cholangiocarcinoma
- Malignant Peripheral Nerve Sheath Tumors
- Myelodysplastic Syndrome (MDS)
- Small Cell Lung Cancer
- Testicular Germ Cell Tumor

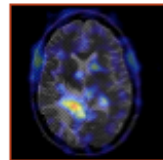
\* *Closed to new accrual*

# TCGA: "No Platform Left Behind"

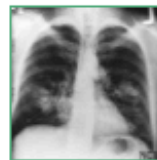


## 36 cancer types

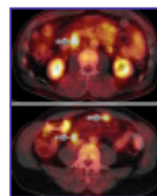
glioblastoma multiforme  
(brain)



squamous carcinoma  
(lung)



serous  
cystadenocarcinoma  
(ovarian)



Etc. Etc. Etc.

Biospecimen Core  
Resource with more  
than 150 Tissue Source  
Sites

5 Cancer Genomic  
Characterization  
Centers

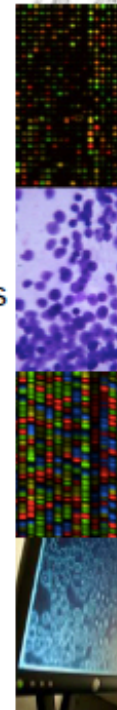
3 Genome  
Sequencing  
Centers

7 Genome Data  
Analysis Centers

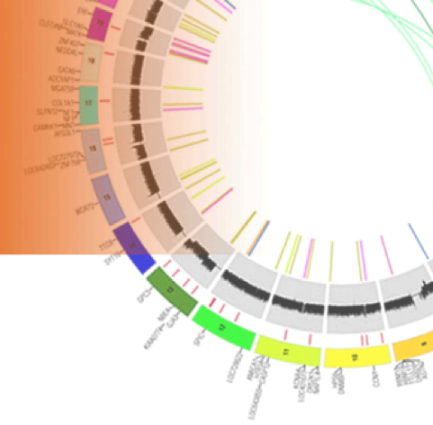
Data Coordinating  
Center

## Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic report/images
- Tissue anatomic site
- Surgical history
- Gene expression/RNA sequence
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence
- RPPA (protein)
- Subset for Mass Spec

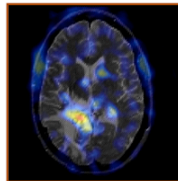


# TCGA: “No Platform Left Behind”

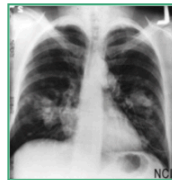


## 36 cancer types

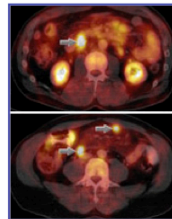
glioblastoma multiforme  
(brain)



squamous carcinoma  
(lung)

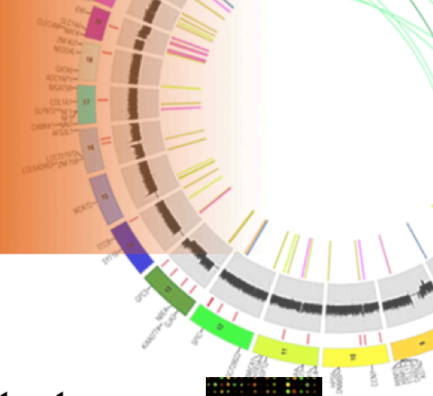


serous  
cystadenocarcinoma  
(ovarian)



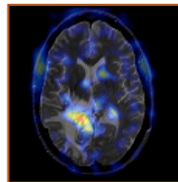
Etc. Etc. Etc.

# TCGA: “No Platform Left Behind”

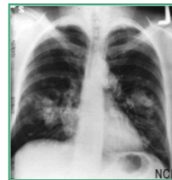


## 36 cancer types

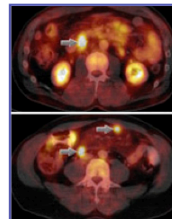
glioblastoma multiforme  
(brain)



squamous carcinoma  
(lung)



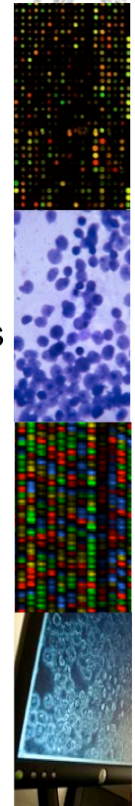
serous  
cystadenocarcinoma  
(ovarian)



Etc. Etc. Etc.

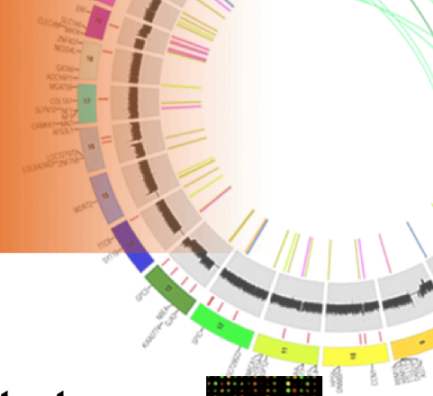
## Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic report/images
- Tissue anatomic site
- Surgical history
- Gene expression/RNA sequence
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence
- RPPA (protein)
- Subset for Mass Spec



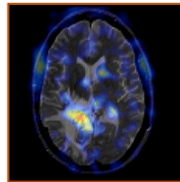


# TCGA: "No Platform Left Behind"

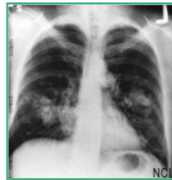


## 36 cancer types

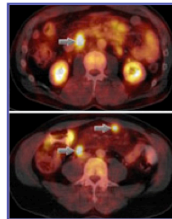
glioblastoma multiforme  
(brain)



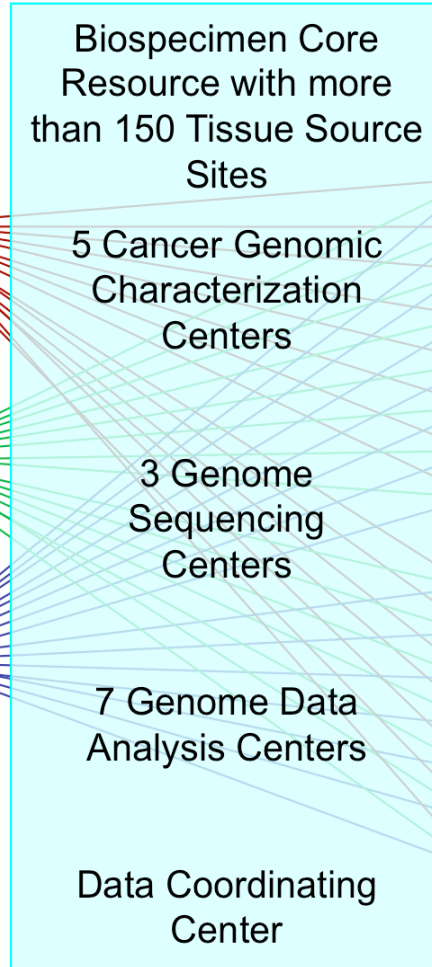
squamous carcinoma  
(lung)



serous  
cystadenocarcinoma  
(ovarian)

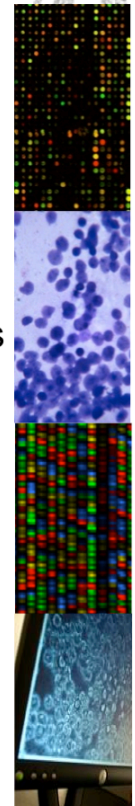


Etc. Etc. Etc.



## Multiple data types

- Clinical diagnosis
- Treatment history
- Histologic diagnosis
- Pathologic report/images
- Tissue anatomic site
- Surgical history
- Gene expression/RNA sequence
- Chromosomal copy number
- Loss of heterozygosity
- Methylation patterns
- miRNA expression
- DNA sequence
- RPPA (protein)
- Subset for Mass Spec



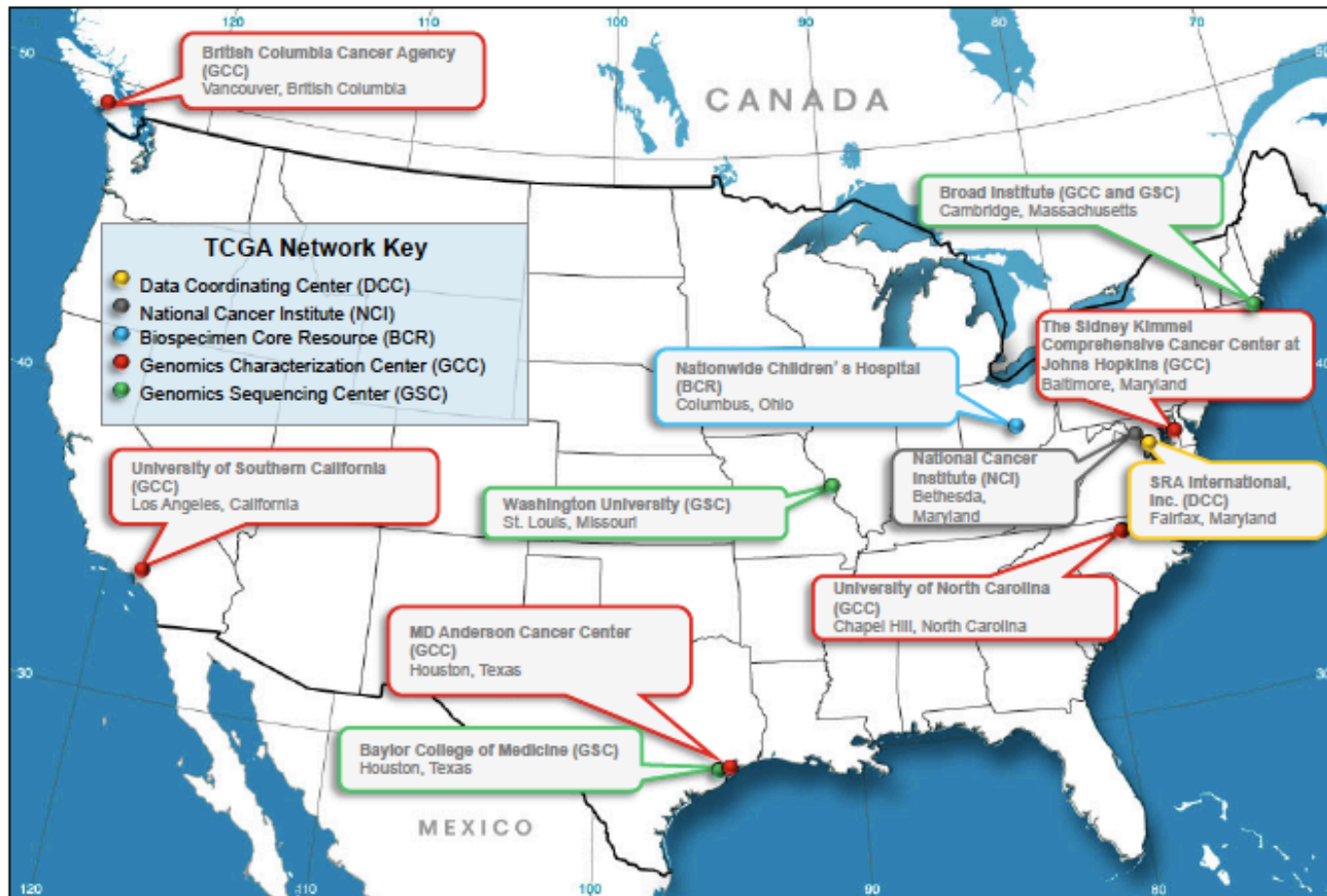
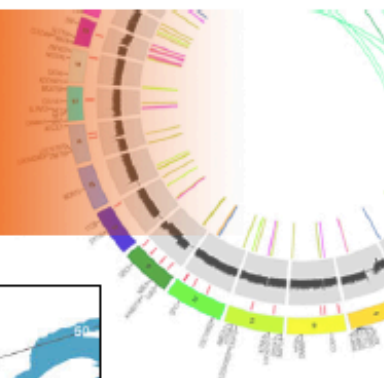


# TCGA: Platforms

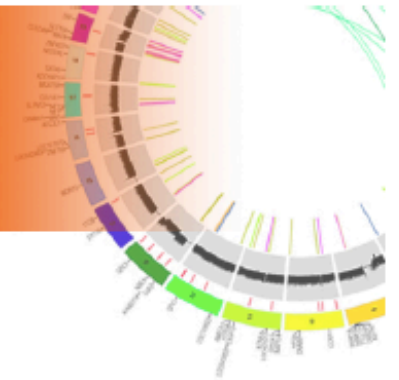


Platform	Pilot (GMB, OV)	Expansion
SNP/copy number	Affymetrix SNP 6.0 Agilent CGH Array Illumina 1M Duo	Affymetrix SNP 6.0 Low-pass sequencing
DNA methylation	Infinium 4500K array	Infinium 4500K array
mRNA	Agilent 244K Array Affymetrix Human Exon Array Affymetrix U133 Array	RNAseq
miRNA	Agilent 8X15K Array	RNAseq
Mutation	600-1000 genes	Illumina exome sequencing Illumina whole genome sequencing (10%)
Proteomics	N/A	Reverse-phase protein array; mass spectrometry

# TCGA Network

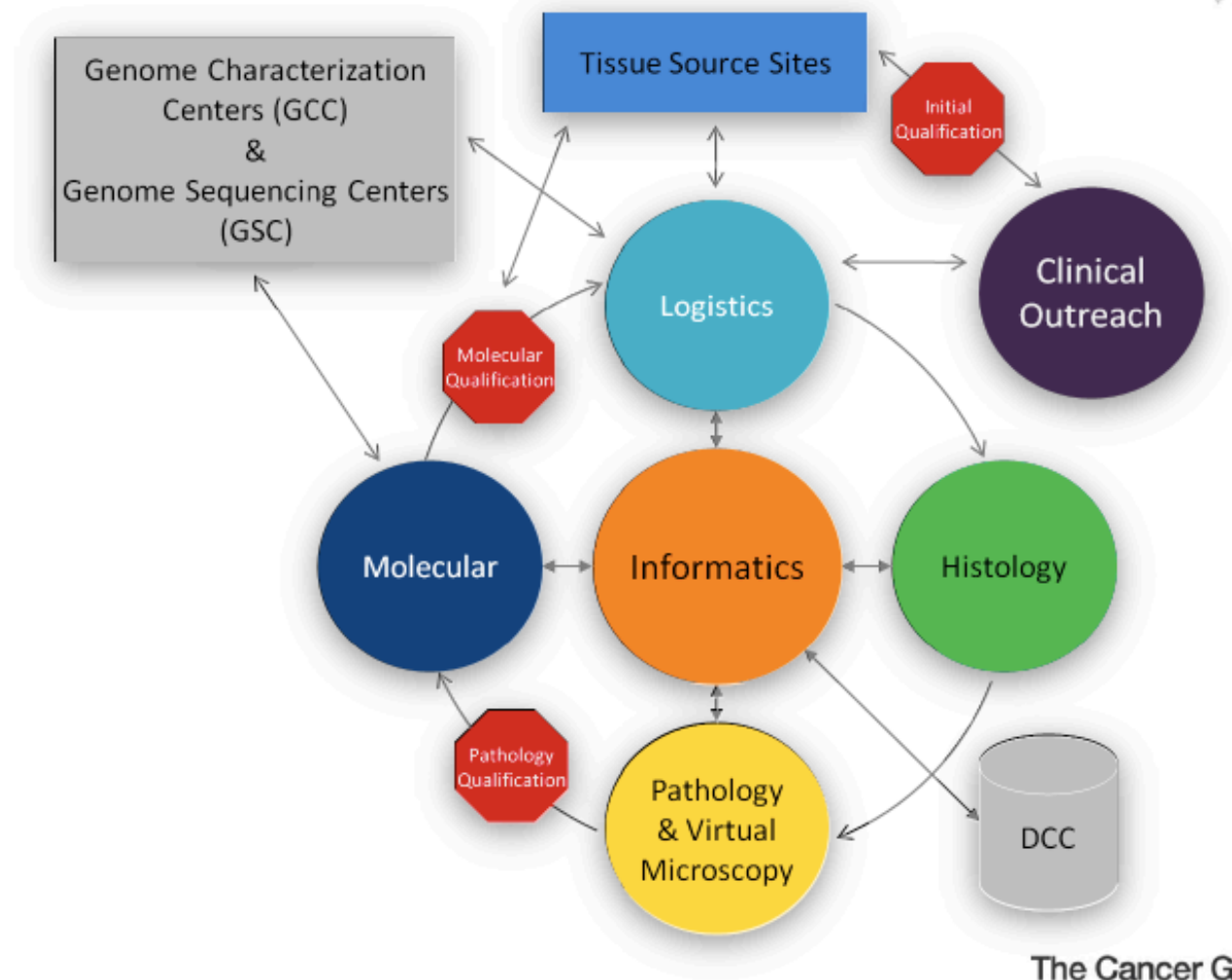
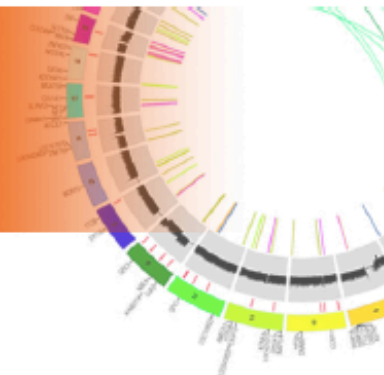


# The BCR Role

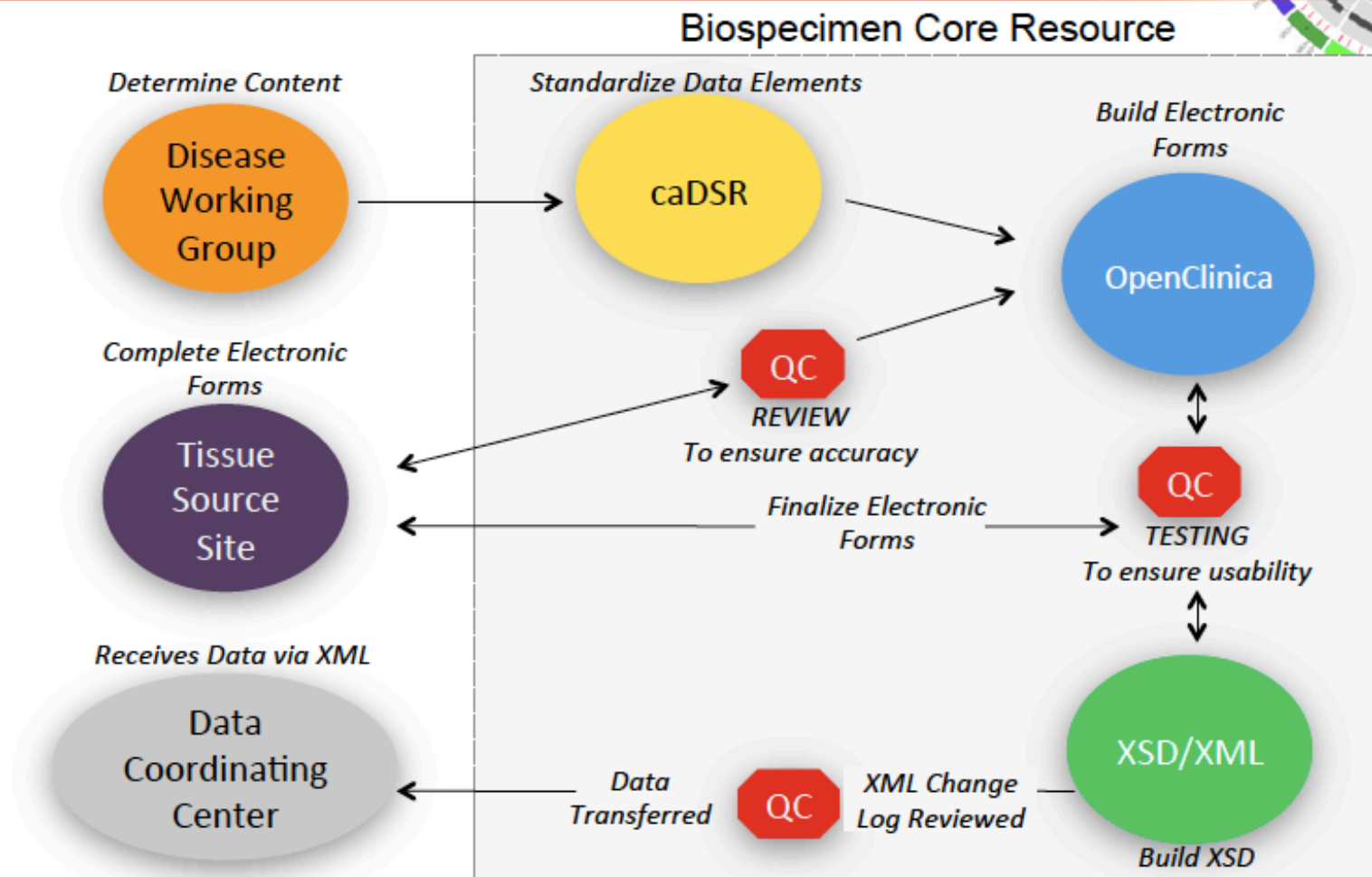


- **Onboarding and Training of Tissue Source Sites (TSS)**
  - IRB/MTA coordination
  - OpenClinica training (web-based clinical data capture system)
  - On-site training and shipment coordination
- **Pathology Validation**
  - Confirmation of acceptable TCGA histology
  - Verification of tumor nuclei and necrosis metrics via BCR pathologists
- **Molecular Processing**
  - Standardized co-isolation of DNA and RNA
  - Quality control, including RNA integrity determination and identity panel
  - Distribution of nucleic acids to Genome Characterization Centers (GCCs) and Genome Sequencing Centers (GSCs)
- **Clinical Data Collection**
  - Development of clinical data forms for TCGA studies
  - Coordination, collection and quality control of clinical data
  - Electronic transfer to Data Coordinating Center (DCC)
  - Assistance with TCGA clinical data analysis

# BCR Pipeline



# TCGA Clinical Data Process



# Clinical Form Development

- **Form Creation**
  - Generic forms are modified by Disease Working Groups for each tumor type
- **Electronic Build**
  - Wire-frame forms are used as a tool to communicate and build studies.
  - Electronic forms are developed and launched in the OpenClinica system
- **Form Completion by TSSs**
  - Tissue Source Sites complete Case Report Forms (CRFs) using OpenClinica
- **Data Review and Quality Control**
  - All forms are reviewed for accuracy by the BCR, who works with TSSs to correct discrepancies.

**Initial Submission Case Quality Control Form**  
Generic

Tissue Source Site (TSS) \_\_\_\_\_ TSS Identifier: \_\_\_\_\_ TSS Unique Patient Identifier: \_\_\_\_\_  
Interviewer Name: \_\_\_\_\_ Interview Date: \_\_\_\_\_

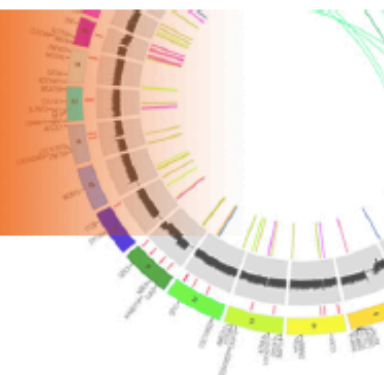
**Tumor Information:** The following sections are to be completed by the TSS Principal Investigator or PI delegate.

Tissue Source Site (TSS) acknowledges that the Biorepository Case Report Form (BCR) may require that the diagnostic slide from a specimen is submitted to the primary diagnostic reporter by the TSS through Anatomic Pathology communication to the BCR laboratory. If the BCR identifies a possible discrepancy, the TSS will be required to report these potential results to the TSS manager, journal report or complete/total form for the quality control program of the TSS to address.

#	Question	Entry Alternatives	Marking Instructions
1	Tumor Identifier	_____	Provide the TSS unique tumor ID. If multiple pieces of tumor are submitted, each tumor needs a unique ID.
2	Diagnosis	<input type="checkbox"/> Tumor Etiology	Indicate the estimated diagnosis of the tumor submitted for TSSA.
3	Other Diagnosis	_____	If diagnosis of tumor submitted for TSSA is not submitted, this question should be skipped.
4	Tumor Type	<input type="checkbox"/> Primary <input type="checkbox"/> (Primary) (resected) (metastatic) (biopsy) (met)	Indicate the type of tumor submitted for TSSA.
5	Anatomic Site of Tumor Biopsy/Specimen	<input type="checkbox"/> Site A-Z (Anatomical) or cell number, if applicable	Indicate the anatomic site of the tumor submitted for TSSA.
6	Month of Cancer Sample Procurement	01 <input type="checkbox"/> 02 <input type="checkbox"/> 03 <input type="checkbox"/> 04 <input type="checkbox"/> 05 <input type="checkbox"/> 06 <input type="checkbox"/> 07 <input type="checkbox"/> 08 <input type="checkbox"/> 09 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/> 12 <input type="checkbox"/>	Indicate the month of the procedure performed to obtain the malignant tissue submitted for TSSA.
7	Day of Cancer Sample Procurement	01 <input type="checkbox"/> 02 <input type="checkbox"/> 03 <input type="checkbox"/> 04 <input type="checkbox"/> 05 <input type="checkbox"/> 06 <input type="checkbox"/> 07 <input type="checkbox"/> 08 <input type="checkbox"/> 09 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/> 12 <input type="checkbox"/> 13 <input type="checkbox"/> 14 <input type="checkbox"/> 15 <input type="checkbox"/> 16 <input type="checkbox"/> 17 <input type="checkbox"/> 18 <input type="checkbox"/> 19 <input type="checkbox"/> 20 <input type="checkbox"/> 21 <input type="checkbox"/> 22 <input type="checkbox"/> 23 <input type="checkbox"/> 24 <input type="checkbox"/> 25 <input type="checkbox"/> 26 <input type="checkbox"/> 27 <input type="checkbox"/> 28 <input type="checkbox"/> 29 <input type="checkbox"/> 30 <input type="checkbox"/> 31 <input type="checkbox"/>	Provide the day of the procedure performed to obtain the malignant tissue submitted for TSSA.
8	Year of Cancer Sample Procurement	_____	Provide the year of the procedure performed to obtain the malignant tissue submitted for TSSA.
9	Method of Cancer Sample Procurement	<input type="checkbox"/> Blood Draw <input type="checkbox"/> Surgical Resection <input type="checkbox"/> Excisional Biopsy <input type="checkbox"/> Other, specify: _____	Indicate the procedure performed to obtain the malignant tissue submitted for TSSA.
10	Other Method of Cancer Sample Procurement	_____	If a procedure performed to obtain the malignant tissue is not included as the provided list, specify the procedure.
11	Country Where Cancer Sample was Procured	_____	Provide the country where the tumor submitted for TSSA was procured.
12	Vessel Used	<input type="checkbox"/> Cryovial <input type="checkbox"/> Cryovial <input type="checkbox"/> Corvette <input type="checkbox"/> Disposable Storage Bag <input type="checkbox"/> Other, specify: _____	Indicate the type of vessel used to transport the tissue to the Biorepository Case Report Form (BCR) for TSSA.
13	Other Vessel Used	_____	
14	Weight of Frozen Tumor	_____ [mg]	Provide the weight of the tumor sample submitted for TSSA.
15	Tumor Nuclei %	_____ [%]	Provide the percentage of nuclei for the sample submitted for TSSA. Check with the BCR to confirm the correct acceptable TSSA metric.
16	Necrosis %	_____ [%]	Provide the percentage of necrosis for the sample submitted for TSSA. Check with the BCR to confirm the correct acceptable TSSA metric.
17	Was sample preserved at site?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Indicate whether the sample submitted to the BCR was preserved at the TSS.
18	Will any slide be sent to the BCR?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Indicate whether a physical copy slide for the sample submitted to the BCR will be shipped with the tumor sample.
19	Will digital slide image be sent to the BCR?	<input type="checkbox"/> Yes <input type="checkbox"/> No	Indicate whether a digital slide image for the sample submitted to the BCR will be shipped with the tumor sample.



# Clinical Data Forms

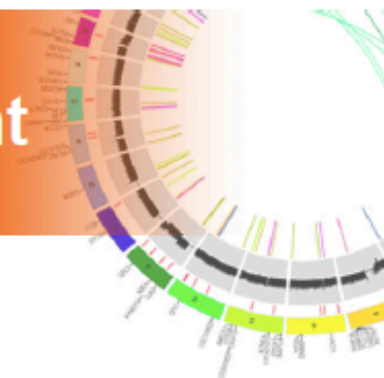


- **Case Quality Control Form** (*required for all cases*)
  - Information about the submitted samples
- **Enrollment Form** (*required for all qualified cases*)
  - Disease-specific information and detailed patient information
- **Follow-up Form(s)** (*required for all living patients*)
  - Updated information not received at time of enrollment
- **Supplemental Treatment Form** (*optional data*)
  - Pharmaceutical Form: Chemotherapy, immunotherapy, targeted therapy, and hormone therapy given for the submitted malignancy
  - Radiation Form: Radiation treatment given for the submitted malignancy
  - Ablation/Embolization Form
- **Other Malignancy Form** (*if applicable*)
  - Captures diagnosis and treatment information related to prior malignancies and synchronous primaries

<http://www.nationwidechildrens.org/biospecimen-core-resource-for-the-cancer-genome-atlas>

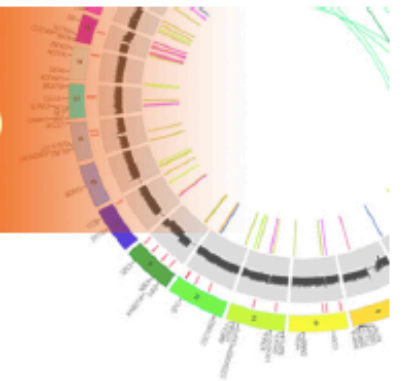


# Minimum Clinical Data Set – Enrollment



1. TSS name/identifier/patient #
2. Primary site of disease
3. Histological type
4. Tumor grade
5. Gender
6. Date of birth (month/year)
7. Prior diagnosis of prior neoplasm (*if yes, another form required*)
8. Date of initial pathologic diagnosis (month/year)
9. Neoadjuvant therapy
10. Tumor stage
11. Date of last contact (month/year)
12. Vital status (month/year)
13. Date of death, if applicable (month/year)
14. New tumor events
15. Tumor-specific elements developed by relevant Disease Working Group

# Minimum Clinical Data Set – Follow-Up



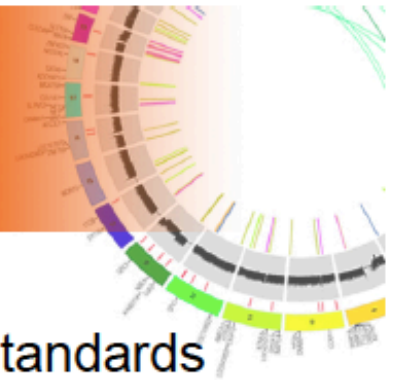
1. Type of treatment incurred
2. Measure of success to treatment
3. Recurrences, progression since last form completed
4. New tumor events
5. Additional surgeries
6. Date of last contact (month/year)
7. Vital status (month/year)
8. Date of Death, if applicable

## Examples of Tumor Specific Questions



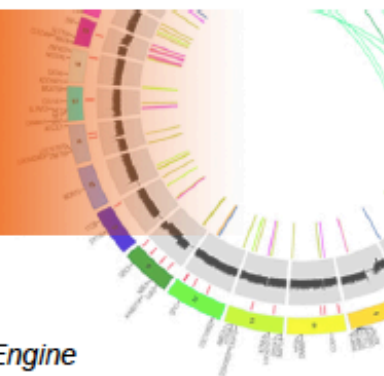
- Smoking status (lung/bladder)
- ER/PR/Her2+ (breast)
- Biochemical recurrence (prostate)
- Residual tumor after surgery (ovarian/sarcoma)
- Prior melanoma diagnosis details (melanoma)
- International Prognostic Index (lymphoma)

# TCGA Data Standardization



- All data elements registered by the Cancer Data Standards Registry (caDSR)
- Common Data Elements (CDEs) – caDSR identifiers are assigned to each TCGA data question
- Standardization allows cross comparison of data elements across TCGA and non-TCGA studies
- Disease site coding standardization according to International Classification of Diseases, version 10 (ICD-10 Coding)
- ICD-O-3 histology coding

# Quality Control of Clinical Data



- OpenClinica Validations
- BCR Validation Engine
- Manual Review

BCR Validation Engine

Site/Identifier	Enrollment	Visit	Validation	Severity	Frequency	Issue	CC Disposition	Notes
05/01/001	05/10/2012	AL-004	Auto Resolved	High	0	BCA 4002		1
05/01/001	05/10/2012	AL-008	Auto Resolved	High	0	BCA 4002		1
05/01/001	05/10/2012	AL-016	Auto Resolved	Low	0	BCA 5481		1
05/01/001	05/10/2012	AL-018	Auto Resolved	Low	0	BCA 5189		1
07/11/001	07/10/2012	T06-2	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-2	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-2	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
07/11/001	07/10/2012	T06-3	Resolved	Medium	0	BCA 4435		1
18/04/001	08/10/2012	T06-3	Not	Medium	0	BCA 4584		1
07/15/001	07/10/2012	T06-3	Not	Medium	0	BCA 4579		1
07/15/001	07/10/2012	T06-3	Not	Medium	0	BCA 4581		1
07/15/001	08/10/2012	T06-3	Resolved	Medium	0	BCA 5173		1
07/15/001	08/10/2012	T06-3	Resolved	Medium	0	BCA 5173		1

Date: dd/mm/yyyy						
TSS: Tissue Source Site Name— Follow Up						
Clinical Outreach Coordinator: Coordinator Name						
SITE	OC ID	FORM	QUESTION	ISSUE	DISCREPANCY NOTE?	RESOLUTION TO ISSUE
XX	###	Follow Up	Enrollment Date of Last Contact: 7/1/09 Follow Up Date of Last Contact: 10/17/04	Follow Up Date of Last Contact should be <b>after</b> Enrollment Date of Last Contact	No	
XX	###	Follow Up	Enrollment Date of Last Contact: 7/1/09 Follow Up Date of Last Contact: 11/20/04	Follow Up Date of Last Contact should be <b>after</b> Enrollment Date of Last Contact	No	

*As the BCR reviews cases, common issues lead to future improvements, which are always centered around standardizing data across forms and disease studies.*

# Transfer of Data



Page 1  
**Consent Form**  
**Blocker (ELCA)**

**Instructions:** This consent form should be completed for each TSS qualified case upon qualification. All information provided on this form should exclude a study from the date of initial diagnosis of contact with the patient (Date of Initial Diagnosis) and "Date of Last Contact" on this. Questions regarding this form should be directed to the Tissue Source Site's primary Clinical Director.

Please note the following definitions for the "Unkown" and "Not Evaluated" answer options:

**Unknown:** This answer option should only be selected if the TSS does not know the information asked in the data being collected. If this answer option is selected for a question that is part of data set, the TSS must complete a discrepancy in its providing a reason why the answer is unknown.

**Not Evaluated:** This answer option should only be selected by the TSS if it believes that the data cannot be obtained. This could be because the test in question was never performed on the patient or the information requested was never disclosed.

Tissue Source Site (TSS) \_\_\_\_\_ TSS Identifier: \_\_\_\_\_ TSS Unique Patient ID \_\_\_\_\_

Completed By (provide name, title, phone/fax) \_\_\_\_\_ Completed Date: \_\_\_\_\_

**General Information**

Question	Entry Information	Working Instructions
1. Back to TSS required per time from the TSS to provide time intervals as indicated for requested date on this form?	Yes <input type="checkbox"/> No <input type="checkbox"/>	For this question, the TSS should provide the date of initial diagnosis of contact with the patient (Date of Initial Diagnosis) and "Date of Last Contact" on this form. If the TSS cannot provide the date of initial diagnosis, the TSS should select "Not Evaluated".
2. Is this a prospective tissue collection?	Yes <input type="checkbox"/> No <input type="checkbox"/>	For this question, the TSS should select "Yes" if the tissue is being collected prospectively. If the tissue is being collected retrospectively, the TSS should select "No".
3. Is this a retrospective tissue collection?	Yes <input type="checkbox"/> No <input type="checkbox"/>	For this question, the TSS should select "Yes" if the tissue is being collected retrospectively. If the tissue is being collected prospectively, the TSS should select "No".

**Additional Information**

Question	Entry Information	Working Instructions
4. Month of Birth	01 <input type="checkbox"/> 02 <input type="checkbox"/> 03 <input type="checkbox"/> 04 <input type="checkbox"/> 05 <input type="checkbox"/> 06 <input type="checkbox"/> 07 <input type="checkbox"/> 08 <input type="checkbox"/> 09 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/> 12 <input type="checkbox"/>	Provide the month of birth for the patient. If the patient's date of birth is unknown, select "Not Evaluated".
5. Day of Birth	01 <input type="checkbox"/> 02 <input type="checkbox"/> 03 <input type="checkbox"/> 04 <input type="checkbox"/> 05 <input type="checkbox"/> 06 <input type="checkbox"/> 07 <input type="checkbox"/> 08 <input type="checkbox"/> 09 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/> 12 <input type="checkbox"/> 13 <input type="checkbox"/> 14 <input type="checkbox"/> 15 <input type="checkbox"/> 16 <input type="checkbox"/> 17 <input type="checkbox"/> 18 <input type="checkbox"/> 19 <input type="checkbox"/> 20 <input type="checkbox"/> 21 <input type="checkbox"/> 22 <input type="checkbox"/> 23 <input type="checkbox"/> 24 <input type="checkbox"/> 25 <input type="checkbox"/> 26 <input type="checkbox"/> 27 <input type="checkbox"/> 28 <input type="checkbox"/> 29 <input type="checkbox"/> 30 <input type="checkbox"/> 31 <input type="checkbox"/>	Provide the day of birth for the patient. If the patient's date of birth is unknown, select "Not Evaluated".
6. Year of Birth	01 <input type="checkbox"/> 02 <input type="checkbox"/> 03 <input type="checkbox"/> 04 <input type="checkbox"/> 05 <input type="checkbox"/> 06 <input type="checkbox"/> 07 <input type="checkbox"/> 08 <input type="checkbox"/> 09 <input type="checkbox"/> 10 <input type="checkbox"/> 11 <input type="checkbox"/> 12 <input type="checkbox"/> 13 <input type="checkbox"/> 14 <input type="checkbox"/> 15 <input type="checkbox"/> 16 <input type="checkbox"/> 17 <input type="checkbox"/> 18 <input type="checkbox"/> 19 <input type="checkbox"/> 20 <input type="checkbox"/> 21 <input type="checkbox"/> 22 <input type="checkbox"/> 23 <input type="checkbox"/> 24 <input type="checkbox"/> 25 <input type="checkbox"/> 26 <input type="checkbox"/> 27 <input type="checkbox"/> 28 <input type="checkbox"/> 29 <input type="checkbox"/> 30 <input type="checkbox"/> 31 <input type="checkbox"/>	Provide the year of birth for the patient. If the patient's date of birth is unknown, select "Not Evaluated".

**Have we heard of your submission. The data has NOT been read. See below for details.**

- [X] This case is a prospective case, it cannot be a retrospective case and vice versa.**

TSS Unique Patient ID:  TSS Identifier:  TSS Unique Patient ID:

**General Information**

**Instructions:** The Consent Form should be completed for each TSS qualified case, upon qualification with the form of initial diagnosis of contact with the patient (Date of Initial Diagnosis) and "Date of Last Contact" on this form. Questions regarding this form should be directed to the Tissue Source Site's primary Clinical Director/Contact at the TSS.

Please note the following definitions for the "Unkown" and "Not Evaluated" answer options on this form:

- Unknown:** This answer option should only be selected if the TSS does not know the information asked in the data being collected. If this answer option is selected for a question that is part of the TSS required data set, the TSS must complete a discrepancy in its providing a reason why the answer is unknown.
- Not Evaluated:** This answer option should only be selected by the TSS if it believes that the information being requested cannot be obtained. This could be because the test in question was never performed on the patient or the information requested was never disclosed.

Page:

TSS Unique Patient ID:  TSS Identifier:  TSS Unique Patient ID:

Has this TSS been read?  Yes  No

**Please Note:**

- Provided time intervals must begin with the date of initial diagnosis (i.e., year in month).
- Only provide a date if you have received permission from the TSS to provide time intervals as a substitute for requested date on this form.

Is this a Prospective Tissue Collection?  Yes  No

**[X] YES** Indicates whether the TSS providing tissue is contracted for prospective tissue collection. If the submitted tissue was selected for the specific purpose of TSS, the tissue has been collected prospectively.

Is this a Retrospective Tissue Collection?  Yes  No

**[X] NO** Indicates whether the TSS providing tissue is contracted for retrospective tissue collection. If the submitted tissue was selected prior to the date the TSS contract was executed, the tissue has been collected retrospectively.

Return to:

2. Electronic CRFs from OpenClinica to BCR Data Warehouse

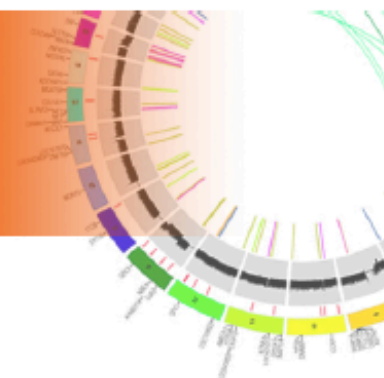


3. Data Warehouse to XML

1. Wireframe Paper CRFs to Electronic OpenClinica CRFs

```
<blca:tissue_prospective_collection_indicator cde="3088492" display_order="1" owner="TSS" preferred_name="Prospective_Collection" procurement_status="Completed" restricted="false" source_system_identifier="674373" tier="1" xsd_ver="2.3">YES</blca:tissue_prospective_collection_indicator>
<blca:tissue_retrospective_collection_indicator cde="3088528" display_order="2" owner="TSS" preferred_name="Retrospective_Collection" procurement_status="Completed" restricted="false" source_system_identifier="674374" tier="1" xsd_ver="2.3">NO</blca:tissue_retrospective_collection_indicator>
<shared:days_to_birth cde="" owner="TSS" precision="day" procurement_status="Completed" tier="1" xsd_ver="1.16">-20885</shared:days_to_birth>
```

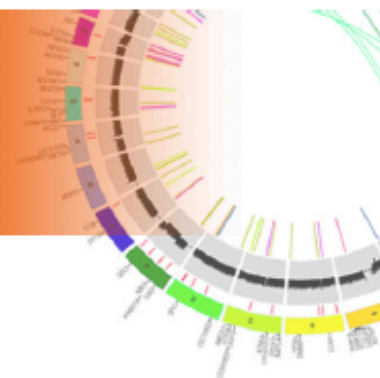
# XML



- EXtensible Markup Language
- Common tool for data transmissions
- Designed to transport and store data
- Allows author to define structure
- Each data element is represented by a string of data tags containing information about the specific element



## Sample XML

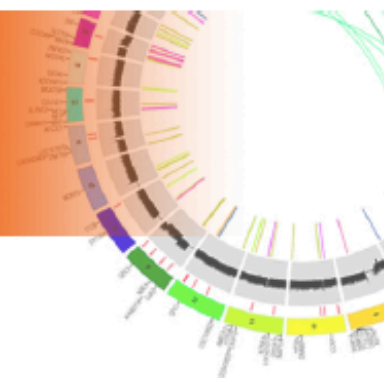


```
<shared:vital_status cde="2939553"  
display_order="22" owner="TSS"  
preferred_name="Vital_Status"  
procurement_status="Completed"  
restricted="false"  
source_system_identifier="674513" tier="1"  
xsd_ver="1.8">DECEASED</shared:vital_status>
```

# Information Provided in the XML

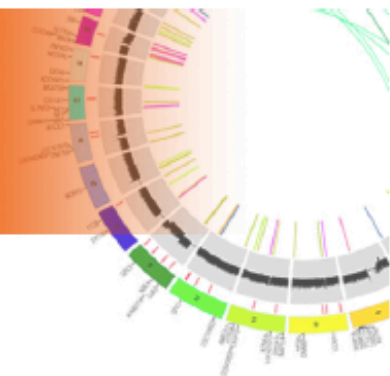
Information Tag	Definition	Example
<b>Data Name</b>	Description of the data element provided and whether it's generic vs. disease specific	<code>&lt;shared:vital_status</code>
<b>CDE Identifier</b>	ID registered at the caDSR	<code>cde="2939553"</code>
<b>Display Order</b>	Order of the element on the CRF	<code>display_order="22"</code>
<b>Owner</b>	Entity that provided the information	<code>owner="TSS"</code>
<b>Preferred Name</b>	Simplified version of "Data Name"	<code>preferred_name="vital_status"</code>
<b>Procurement Status</b>	Collection status of the element	<code>procurement_status="completed"</code>
<b>Restricted</b>	Indicates restricted information	<code>restricted="false"</code>
<b>Source System ID</b>	ID used by OpenClinica	<code>source_system_id="674513"</code>
<b>Tier</b>	Indicates whether the element was required	<code>tier="1"</code>
<b>XSD Version</b>	Version of XSD used to generate XML	<code>xsd_ver="1.8"</code>
<b>Answer Provided by TSS</b>	Information provided by the TSS for this case	<code>&gt;DECEASED&lt;/shared:vital_status&gt;</code>

# Biospecimen XML vs. Clinical XML



- The BCR generates two XML files
- Biospecimen XML
  - Uploads within 24 hours of nucleic acids shipping to GCCs and GSCs
  - BCR Quality Metrics
  - Select Data from Case Quality Control Form
- Clinical XML
  - Uploaded once enrollment form data has passed BCR quality control
  - Select Data from Case Quality Control Form
  - Enrollment Form Data
  - Follow-up Form(s) Data
  - Treatment Form(s) Data

# Biotabs

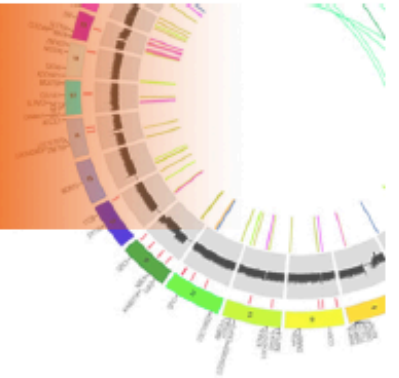


- DCC parses XML uploaded by the BCR
- Data extracted, ‘flattened,’ and put into .txt files
- Can be opened as a tab-delimited file in Excel
- Some data loss
- Some data not required for all tumor types [Not Available]

bcr_patient_barcode	age_at_initial_pathologic_diagnosis	age_began_smoking_in_years	ajcc_cancer_metastasis_stage_code	ajcc_cancer_staging_handbook_edition
TCGA-BL-A0C8	73	13	MO [Not Available] NX Stage I T1	Dome a6003b1c-56a9-430a-a5e2-b70af3f81bdb [Not Available] [Not Available] High Grade
TCGA-BL-A13I	57	27	MO 7th NO Stage III T3	Bladder, NOS ebc094a6-cdc9-4c02-b8ec-a181b25a364d [Not Available] [Not Available] High G
TCGA-BL-A13J	65	25	MO [Not Available] N2 Stage IV T4	Wall, Posterior 556fcb08-172a-4af1-8822-ac036c8d68e8 [Not Available] [Not Available]
TCGA-BL-A3JM	62	14	MO 7th NO Stage III T3	Wall, Lateral 0F4FB265-B19C-4A85-97B3-779429E6CCDD [Not Available] [Not Available] High G
TCGA-BT-A0S7	75	[Not Available]	MX 6th NO Stage III T4a	Bladder, NOS 68bc786e-07c7-4964-bc88-5e2a8633626a [Not Available] [Not Available]
TCGA-BT-A0YX	70	[Not Available]	MO [Not Available] NO Stage III T3b	Bladder, NOS 22473ad4-df39-4716-b558-3c1409d5fa0e pT3b Extensive
TCGA-BT-A20J	75	[Not Available]	MX 5th NO Stage II T2b	Wall, Posterior 74b1e72d-fdd0-4d84-8858-3bb32d58ff7e pT2b [Not Available] High G
TCGA-BT-A20N	72	[Not Available]	MX 6th NO Stage III T3a	Trigone ebedab7c-dd7c-448e-8324-78c86f13be8b pT3a [Not Available] High Grade
TCGA-BT-A20O	75	[Not Available]	MX 6th NO Stage III T3a	Wall, Lateral c0c7fbc7-b9cb-4c65-941b-cd7d581c54cc pT3a [Not Available] High G
TCGA-BT-A20P	81	[Not Available]	MO 6th NO Stage III T3a	[Not Available] f48937ed-c294-41f0-8872-7010391167a9 pT3a [Not Available] High G
TCGA-BT-A20Q	73	[Not Available]	MO 6th N2 Stage IV T3b	[Not Available] 919deaac-ead5-4bdc-9fa0-b5cd8429b44f pT3b [Not Available] High G

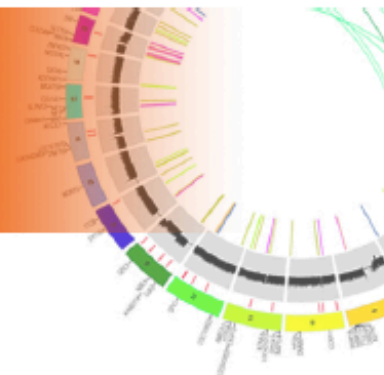
**NOTE:** The BCR is currently working to improve the column headers on the TCGA Data Portal “biotab” files. The goal is to designate a single column header, across studies and forms, for all questions with the same meaning.

# TCGA Public-Access Data



- Data are de-identified
- Molecular platforms:
  - SNP/copy number – level 3 SNP data
  - DNA methylation
  - mRNA
  - miRNA
  - Mutation – level 3 sequencing data
  - Proteomics
- Clinical data

# TCGA Public-Access Data

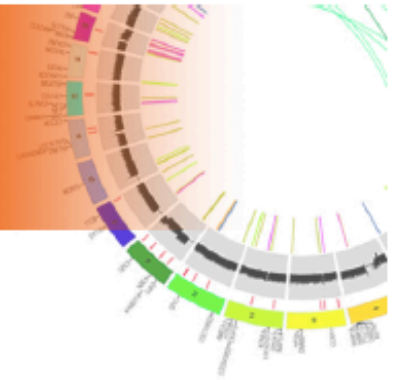


- Redacted pathology reports
- Aperio whole slide images
  - Frozen section of tissue used for extractions
  - Formalin-fixed, paraffin-embedded H&E from a representative diagnostic block
- MSI testing (level 1, 2, and 3)
  - COAD and READ
  - STAD in the future
- HPV testing
  - Currently validating assay and will test HNSC cases
  - CESC in the future

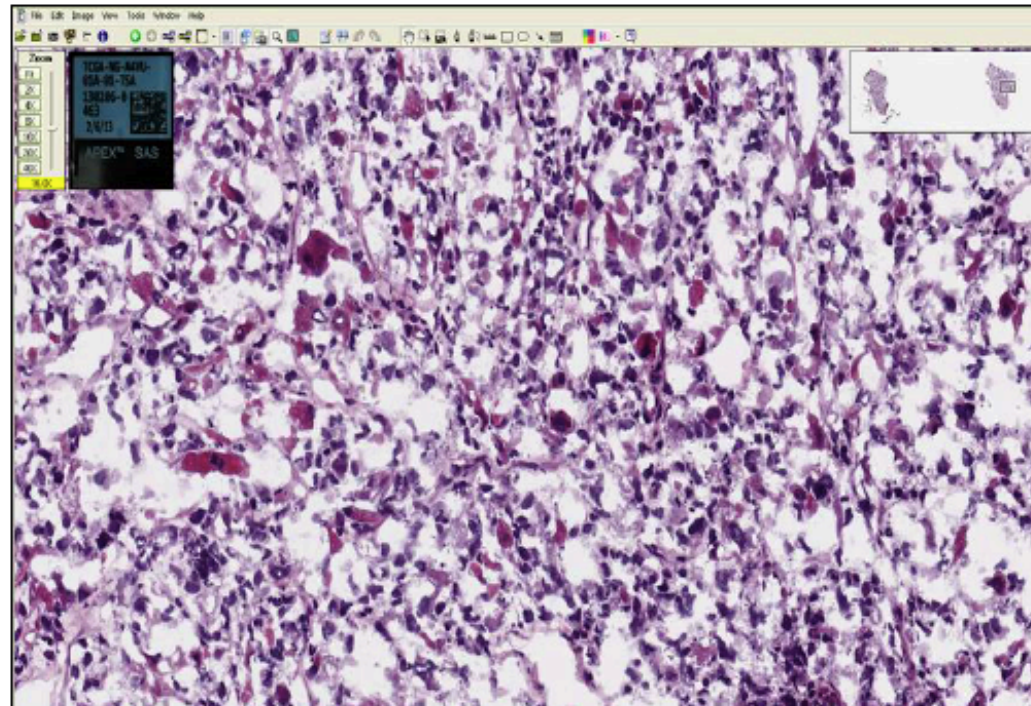




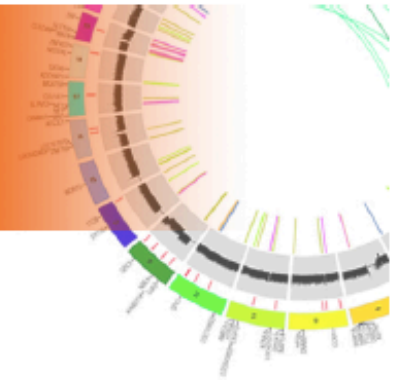
# TCGA Public-Access Data



- Whole slide frozen sections 40X images
- Whole slide FFPE Diagnostic 40X images
- Aperio .svs file format
- JPEG compression
- Quality factor of 30

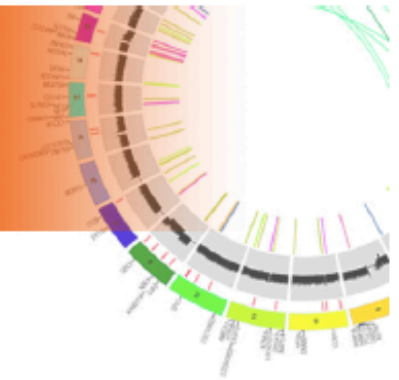


# TCGA Controlled-Access Data



- Data Access Request is required
- Contains data that are unique to an individual
  - Individual germline variant data (SNP .cel files)
  - Primary sequence data - available at the Cancer Genomics Hub (<https://cghub.ucsc.edu/>)
  - Exon array data (applies to the Glioblastoma Multiforme and Ovarian Serous Cystadenocarcinoma projects only)

# Annotations Manager



- For when things don't exactly fit
- Example annotations include:
  - Neoadjuvant therapy
  - History of prior malignancy
  - Subject withdrew consent
  - Synchronous malignancy
  - Redactions

The Cancer Genome Atlas Data Portal

understanding genomics to improve cancer care

TCGA Home | Contact Us | Feedback

Home | Query the Data | Download Data | Tools | About the Data | Publication Guidelines

Home » Tools » Annotations Manager

In This Section

### Annotations Manager

TCGA Annotations

Welcome to TCGA Annotations

Please log into get access to restricted data

Search Annotations

Search Options:  By Filtering  By ID

Disease:

Barcode of Items:

Item Type:

Annotation Classification:

Annotation Category:

Keyword:

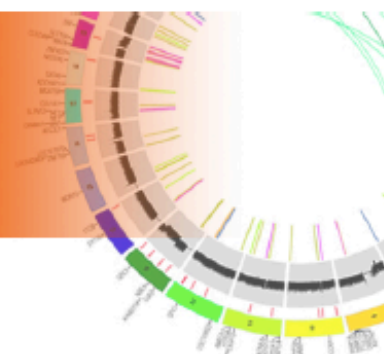
Annotation:

Exact Match

Include Prescribed Annotations

Search Reset Cancel

# TCGA Number Decoded



TCGA-BL-A123-01A\*-11D\*-A123\*-02

**PROJECT NAME**

The first part of the number represents the current project Name. Currently TCGA is the only project.

**TISSUE SOURCE SITE (TSS)**

This will be a numeric/alpha-numeric ID unique to each cancer-type at each TSS.

**PATIENT IDENTIFIER**

This ID is unique to each patient and is associated with a site ID

**SAMPLE IDENTIFIER**

The sample ID represents both the sample type and the sample vial, as follows:

Sample type:

- Two digit # 01-29 (see details below)
  - #s 01-09 = tumor types
  - #s 10-19 = normal types
  - #s 20-29 = control samples

Vial Identifier:

An alphabetic character represents the vial of a specific specimen. The values range from A-Z.

**Sample ID Details**

01 Solid Tumor	12 Buccal Smear
10 Normal Blood	20 Cell Line
11 Normal Tissue	

**PLATE IDENTIFIER**

Identifies individual plates associated with the sample.

**CENTER IDENTIFIER**

Identifies each of the Genomic Sequencing Centers (GSCs) and Cancer Genome Characterization Centers (CGCCs)

**PORTION IDENTIFIER**

The portion ID consists of a portion code and an analyte code.

- Portion Code: A two digit number identifying the portion
- Analyte Code: Alphabetic code representing an analyte type (i.e. D for DNA). (see details below)

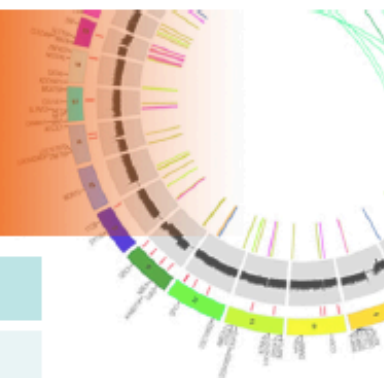
**Analyte Code Details**

R RNA	D DNA
T Total RNA	W WGA DNA (Qiagen)

**IMPORTANT NOTES\***

- \* **SLIDES:** When the slides are cut an additional number will be added to after the Vial ID, to determine whether the slide is a top slide (TS1), bottom slide (BS1), or a middle slide (MS1).
- \* **ANALYTE CODE:** The analyte code will not be added until the specimen has passed the BCR pathology review.
- \* **PLATE ID:** The plate ID and the center ID will not be added until the specimen has been through the Molecular workflow.

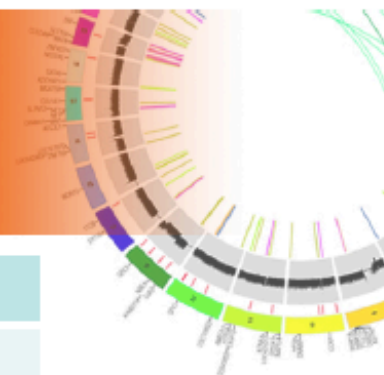
# Tumor Codes



Tumor	Code
Acute Myeloid Leukemia	LAML
Bladder Urothelial Carcinoma	BLCA
Breast Invasive Carcinoma	BRCA
Cervical Carcinoma	CESC
Colon Adenocarcinoma	COAD
Rectal Adenocarcinoma	READ
Diffuse Large B-Cell Lymphoma	DLBC
Endometrial Carcinoma	UCEC
Esophageal Carcinoma	ESCA
Glioblastoma Multiforme	GBM
Head and Neck Carcinoma	HNSC
Hepatocellular Carcinoma	LIHC



# Tumor Codes



Tumor	Code
Kidney Clear Cell Carcinoma	KIRC
Kidney Papillary Carcinoma	KIRP
Lower Grade Glioma	LGG
Lung Adenocarcinoma	LUAD
Lung Squamous Cell Carcinoma	LUSC
Melanoma	SKCM
Ovarian Serous Carcinoma	OV
Pancreatic Adenocarcinoma	PAAD
Prostate Adenocarcinoma	PRAD
Stomach Adenocarcinoma	STAD
Sarcoma	SARC
Thyroid Papillary Carcinoma	THCA



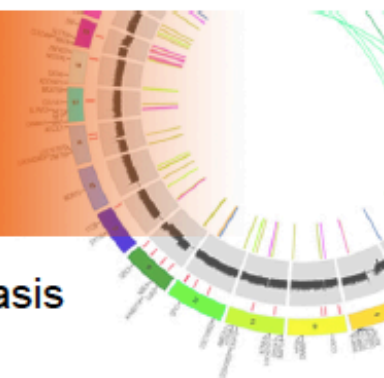
# TCGA Publication Policy



- The TCGA program has established the following policy to clarify freedom of TCGA and non-TCGA users to publish findings using TCGA data. There are no limitations on publications containing analyses using any TCGA data set if the data set meets one of the following three freedom-to-publish criteria: A marker paper has been published on that tumor type; or
  - 18 months after 100 cases of a given tumor type have shipped from BCR to characterization and sequencing centers; or
  - The author receives specific approval from the TCGA Publication Committee in consultation with appropriate disease-specific analysis group(s).
  - Specifically, the status of each tumor dataset is available below. If you have questions, do not hesitate to contact [tcga@mail.nih.gov](mailto:tcga@mail.nih.gov).

TCGA Data Use Policy and Publication Guidelines promote the responsible use of TCGA data sets. All investigators, and their institutions, seeking access and use of TCGA data must acknowledge their agreement with TCGA policies and procedures, which are described in the TCGA Data Portal [Data Use Certification](#).

# Accessing TCGA Data



New data derived from TCGA analyses is deposited on a regular basis into the TCGA Data Portal. This process is controlled by the Data Coordinating Center (DCC).

<http://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp>

Home | Query the Data | Download Data | Tools | About the Data

**TCGA Data Portal Overview**

We provide 3 ways to download data: The Cancer Genome Atlas (TCGA) Data Portal provides a platform for researchers to search, download, and analyze data sets generated by TCGA. It contains clinical information, genomic characterization data, and high-throughput sequencing analysis of the tumor genomes.

[Query the Data](#) | [Download Data](#)

Search summarized data for genes, patients and pathways | Choose from three ways to download data

Available Cancer Types	# Patients with Samples	# Downloadable Tumor Samples	Date Last Updated (mm/dd/yy)
Acute Myeloid Leukemia [AML]	202	200	02/17/11
Bladder Urothelial Carcinoma [BLCA]	26	12	04/15/11
Brain Lower Grade Glioma [LGG]	30	0	04/14/11

**Announcements**

08/20/2010 - Data Access Matrix Web Service now available

A REST-style web service interface is now available for the Data Access Matrix application. Please see the [Web Service User Guide](#) for detailed instructions and further information.

08/20/2010 - FTP/SFTP protocols deprecated

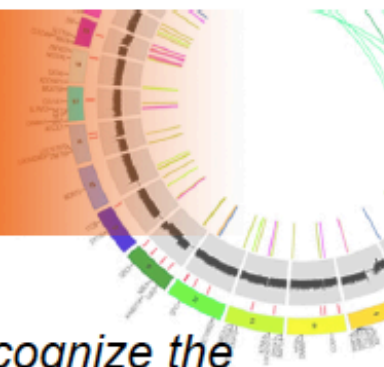
Please see [this wiki news post](#) for more information.

[See all announcements](#)

**More TCGA Information**

More information about The Cancer Genome Atlas

# Participate in TCGA



*We welcome additional collaborators at any time as we recognize the need to involve teams from around the world to acquire the samples and data necessary for a program of this size and scope.*

## **The Cancer Genome Atlas Program Office**

National Cancer Institute at NIH

31 Center Drive

Bldg. 31, Suite 3A20

Bethesda, MD 20892

Phone: (301) 594-9831

E-mail: [tcga@mail.nih.gov](mailto:tcga@mail.nih.gov)