



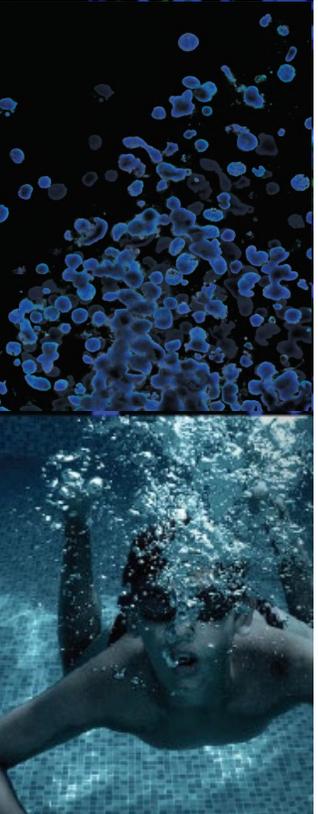
Navigating the HCMI Searchable Catalog and Accessing HCMI Data at NCI's Genomic Data Commons

Eva Tonsing-Carter, PhD
Scientific Program Manager
NCI Center for Cancer Genomics

Bill Wysocki, PhD
Lead for User and Data Services
University of Chicago

James Clinton, PhD
Lead Scientist
ATCC

Credible Leads to Incredible™



About ATCC

- Founded in 1925, ATCC is a non-profit organization with HQ in Manassas, VA, and an R&D and Services center in Gaithersburg, MD
- World's largest, most diverse biological materials and information resource for cell culture – the “*gold standard*”
- Innovative R&D company featuring gene editing, differentiated stem cells, advanced models
- cGMP biorepository
- Partner with government, industry, and academia
- Leading global supplier of authenticated cell lines, viral and microbial standards
- Sales and distribution in 150 countries, 19 international distributors
- Talented team of 450+ employees, over one-third with advanced degrees

Our speakers



Eva Tonsing-Carter, PhD

Scientific Program Manager

National Cancer Institute's Center for Cancer Genomics



NATIONAL CANCER INSTITUTE

Center for Cancer Genomics



Bill Wysocki, PhD

Lead for User and Data Services

Genomic Data Commons in the Center for Translational
Data Science (University of Chicago)



NATIONAL CANCER INSTITUTE

Genomic Data Commons

Human Cancer Models Initiative (HCMI): Navigating the HCMI Searchable Catalog and Accessing HCMI Data at NCI's Genomic Data Commons

Eva Tonsing-Carter, Ph.D.

Health Scientist Administrator (Program Officer)

Center for Cancer Genomics

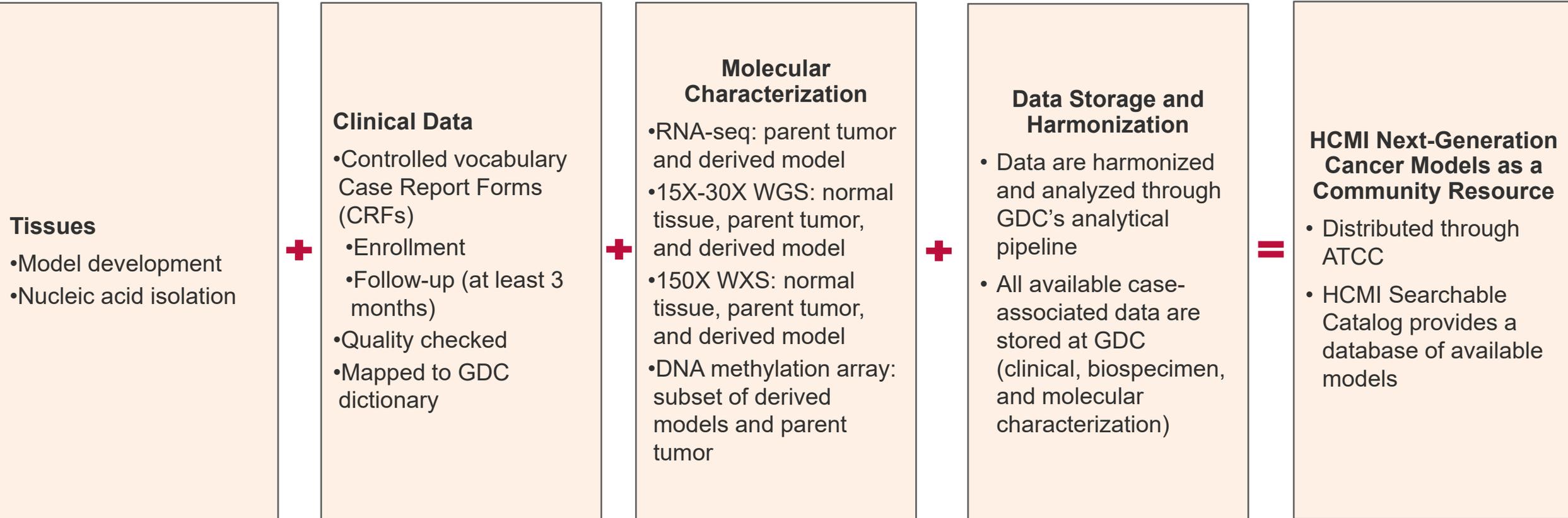
National Cancer Institute

Human Cancer Models Initiative (HCMI) Program

- International consortium established by:
 - National Cancer Institute (NCI)
 - Cancer Research UK (CRUK)
 - Wellcome Sanger Institute (WSI)
 - Hubrecht Organoid Technology (HUB)
- The consortium's **objective** is to provide next-generation cancer models derived from patient tissues as a scientific community resource.
 - provide models derived from a range of tumor types
 - models are associated with clinical and molecular characterization data
 - models along with associated data will be broadly available to the scientific community



NCI-supported Cancer Model Development Pipeline



<https://www.cancer.gov/ccg/research/functional-genomics/hcmi/about/cancer-model-development>

HCMI Resources

- HCMI Searchable Catalog: <https://hcmi-searchable-catalog.nci.nih.gov>
- Office of Cancer Genomics HCMI Resources: <https://www.cancer.gov/ccg/research/functional-genomics/hcmi/using-hcmi>
 - Cancer-specific Case Report Forms
 - Informed Consent Template
 - HCMI Searchable Catalog User Guide
- Model Distributor: www.atcc.org/hcmi
 - HCMI models, protocols, and standardized reagents
- Genomic Data Commons (GDC): <https://portal.gdc.cancer.gov/projects/HCMI-CMDC>
 - Clinical, biospecimen, and molecular characterization data

HCMI Resources – Case Report Forms

Case Report Forms

Human Cancer Models Initiative's (HCMI) cancer type specific case report forms (CRFs) have been developed through collaborations with international clinical experts and the clinical data elements have been standardized through the Cancer Data Standards Registry and Repository (caDSR). As the tumor types modeled through the HCMI are constantly updated, be sure to check back regularly for additional or updated CRFs.

Please note that in some cases, it is possible to collect tissues from different tumor sites (e.g., primary, metastatic, or recurrent) from the same patient for model development. New CRFs capture information about multiple models developed from the same patient within a single form. The multiple model CRFs are designated with "-multi" within the file name. Versioning is used if there are any subsequent edits to the CRFs.

The cancer types that are not highly represented in the HCMI models are considered "rare" by HCMI, however, they may not be considered "rare" based on their representation among cancer incidence rates. The "Rare Cancer CRFs" are being used to capture clinical data for these cancer types: Bladder Cancer, Chordoma, Epithelioid Sarcoma, Extrahepatic Cholangiocarcinoma, Gallbladder Cancer, Spindle Cell Sarcoma, Small Intestine Cancer, and Thyroid Cancer.

Enrollment Forms

The enrollment CRFs data are collected at the time of patient index date, which includes initial pathologic diagnosis, sample procurement, or first patient visit.

- Ampulla of Vater-multi Enrollment CRF
- Brain-multi Enrollment CRF
- Breast-multi Enrollment CRF
- Colorectal-multi Enrollment CRF
- Endometrium-multi Enrollment CRF
- Esophageal-multi Enrollment CRF
- Ewing Enrollment CRF
- Glioblastoma-multi Enrollment CRF
- Head and Neck-multi Enrollment CRF
- HCC and Intrahepatic-multi Enrollment CRF
- Kidney-multi Enrollment CRF
- Lung-multi Enrollment CRF
- Melanoma-multi Enrollment CRF
- Mesothelioma-multi Enrollment CRF
- Neuroblastoma Enrollment CRF
- Neuroendocrine Enrollment Suppl CRF
- Osteosarcoma Enrollment CRF
- Ovarian-multi Enrollment CRF
- Pancreas-multi Enrollment CRF
- Pediatric Liver Enrollment CRF
- Prostate-multi Enrollment CRF
- Rare-multi Enrollment CRF
- Rhabdomyosarcoma-multi Enrollment CRF
- Stomach-multi Enrollment CRF
- Wilms Enrollment CRF

Follow-up Forms

The follow-up CRF data are collected at least 3 months after enrollment data collection following model establishment.

V1.0

Enrollment: Brain

Tissue Source Site (TSS) Name: _____ HCMI Identifier (ID3): _____
 Completed By: _____ Completion Date (MM/DD/YYYY): _____

Form Notes: An Enrollment Form should be completed for each HCMI case upon qualification notice from Leidos. All information provided on this form should include activity from the Date of Initial Pathologic Diagnosis to the most recent Date of Last Contact with the patient. This form should be used for the following Brain Cancers: Embryonal Tumor, Medulloblastoma, Diffuse Midline Glioma, and Lower Grade Glioma.

Question	Question Text	Data Entry Options	CDE ID	Instruction Text
1	ID2	_____	2003301	Provide the patient's ID2 (this ID will only be used by IMS for internal quality control).
2	ID3	_____	5845012	Provide the HCMI-specific anonymized ID (ID3).
3	Index date	<input type="checkbox"/> Initial pathologic diagnosis <input type="checkbox"/> Sample procurement <input type="checkbox"/> First patient visit	6154722	Select the reference date used to calculate time intervals (e.g. days to treatment). <i>Date of initial pathologic diagnosis is the HCMI standard and should be used unless it is unavailable. If an alternative index date is used, indicate it here and use it for all interval calculations.</i>
Patient Information				
4	Number of days from index date to date of last contact	_____	3008273	Provide the number of days from the index date to the date of last contact.
5	Patient age on index date	_____	6379572	Provide the age (in days) of the patient on the index date. <i>Note: If the patient's age is greater than 32,872 days (90 years), please enter 32,872.</i>
6	Gender	<input type="checkbox"/> Male <input type="checkbox"/> Female <input type="checkbox"/> Unspecified	2200604	Provide the patient's gender using the defined categories. <i>Identification of gender is based upon self-report and may come from a form, questionnaire, interview, etc.</i>
7	Height	_____	649	Provide the patient's height, in centimeters.
8	Weight	_____	651	Provide the patient's weight, in kilograms.
9	Body mass index (BMI)	_____	2006410	If the patient's height and weight are not collected, provide the patient's body mass index (BMI).
10	Race	<input type="checkbox"/> American Indian or Alaska Native <input type="checkbox"/> Asian <input type="checkbox"/> Black or African American <input type="checkbox"/> Native Hawaiian or other Pacific Islander <input type="checkbox"/> White <input type="checkbox"/> Unknown <input type="checkbox"/> Not reported	2192199	Provide the patient's race using the defined categories. American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment. Asian: A person having origins in any of the peoples of the Far East, Southeast Asia, or in the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam. Black or African American: A person having origins in any of the black racial groups of Africa. Native Hawaiian or other Pacific Islander: A person having origins on any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Island. White: A person having origins in any of the original peoples of Europe, the Middle East, or North Africa.

Page 1 of 25

HCMI Searchable Catalog Resource

- **Objective: develop a public-facing online catalog to query and identify HCMI next-generation cancer models available to the research community**
 - The HCMI Searchable Catalog contains model information (e.g. model type), clinical data (e.g. clinical diagnosis, primary site, etc.), distributor information (e.g. images), and molecular characterization data information.
- Developed by OICR (NCI-sponsored FNLCR managed subcontractor)
- Work in progress and will contain model data from all consortium members
- HCMI Searchable Catalog: <https://hcmi-searchable-catalog.nci.nih.gov>



Landing Page: Includes All Available Models to Query

Human Cancer Models Initiative
Searchable Catalog

Use the filter panel on the left to customize your model search.

[SHARE](#) [VIEW LIST](#)

Models By Primary Site

26 Total

Has Multiple Models

2D Versus 3D Growth

Most Frequently Mutated Genes

Showing 1 - 20 of 295 models

Include 290 unexpanded models
 COLUMNS
EXPORT

Name	Primary Site	Clinical Tumor Diagnosis	Tissue Status	Age At Acquisition (Years)	Age At Diagnosis (Years)	Has Multiple Models	Expansion Status	# Mutated Genes	# Research Somatic Variants	# Clinical Variants	# Histo-Pathological Biomarkers
HCM-BROD-0648-C71	Brain	Glioblastoma	Recurrent	68	63	No	EXPANDED	5326	7110	0	3
HCM-BROD-0227-C43	Skin	Melanoma	Metastasis	40	40	No	EXPANDED	3075	4187	0	0
HCM-SANG-0288-C18	Colon	Colorectal cancer	Primary	75		No	EXPANDED	3228	3908	0	0
HCM-BROD-0569-C43	Skin	Melanoma	Metastasis	79	78	No	EXPANDED	2886	3802	0	1
HCM-CSHL-0426-C18	Colon	Colorectal cancer	Primary	73	72	No	EXPANDED	2701	3183	0	0
HCM-SANG-0273-C18	Colon	Colorectal cancer	Primary	78		No	EXPANDED	2597	2991	0	0
HCM-BROD-0027-C34	Bronchus and lung	Lung cancer	Metastasis	66	65	No	EXPANDED	2313	2868	0	0
HCM-CSHL-0459-C17	Small intestine	Rare cancers	Primary	57	57	No	EXPANDED	2426	2793	0	5
HCM-BROD-0223-C43	Skin	Melanoma	Metastasis	74	73	No	EXPANDED	2187	2679	0	0
HCM-SANG-0282-C18	Colon	Colorectal cancer	Primary	85		No	EXPANDED	2313	2636	0	0
HCM-BROD-0106-C71	Brain	Glioblastoma	Recurrent	56	52	No	EXPANDED	2122	2333	0	3
HCM-SANG-0276-C18	Colon	Colorectal cancer	Primary	78		No	EXPANDED	1768	1976	0	0
HCM-BROD-0334-C43	Skin	Melanoma	Metastasis	72	70	No	EXPANDED	1619	1939	0	1
HCM-CSHL-0174-C22	Intrahepatic bile duct	Intrahepatic bile duct canc	Primary	64	64	No	EXPANDED	1568	1713	0	0
HCM-BROD-0702-C43	Skin	Melanoma	Metastasis	70	69	No	EXPANDED	1367	1658	0	1
HCM-CSHL-0317-C18	Colon	Colorectal cancer	Primary	65	64	No	EXPANDED	1502	1639	0	0
HCM-CSHL-0729-C18	Colon	Colorectal cancer	Primary	60	60	No	EXPANDED	1433	1572	0	6
HCM-BROD-0025-C16	Stomach	Stomach cancer	Primary	74	73	No	EXPANDED	1330	1422	0	0
HCM-BROD-0681-C71	Brain	Glioblastoma	Recurrent	52	50	No	EXPANDED	1096	1166	0	3
HCM-BROD-0679-C43	Skin	Melanoma	Metastasis	69	68	No	EXPANDED	765	878	0	1

Updated: July 31, 2023

Search By Model Name

Search By Altered Gene(s)

Search By Research Somatic Variant

Primary Site

- Pancreas 51
- Brain 50
- Colon 47
- Esophagus 36
- 22 More

Research Somatic Variant Type

Consequence

Model Type

Has Multiple Models

Acquisition Site

Clinical Tumor Diagnosis

Clinical Stage Grouping

Histological Subtype

Histological Grade

Age At Diagnosis (Years)

Gender

Available Molecular Characterizations

Neoadjuvant Therapy

Chemotherapeutic Drug List Available

Licensing Required For Commercial Use

Race

Landing Page: Catalog Features Highlights

Human Cancer Models Initiative

Searchable Catalog

Search By Model Name
 e.g. HCM-BROD-0051-C64, ...

Search By Altered Gene(s)
 e.g. BRAF, EWSR, ...

Search By Research Somatic Variant
 e.g. BRAF V600E, IDH1 R132H, ...

Primary Site
 Pancreas (51)
 Brain (50)
 Colon (47)
 Esophagus (36)
 22 More

Research Somatic Variant Type

Consequence

Model Type

Has Multiple Models

Acquisition Site

Clinical Tumor Diagnosis

Clinical Stage Grouping

Histological Subtype

Histological Grade

Age At Diagnosis (Years)

Gender

Available Molecular Characterizations

Neoadjuvant Therapy

Chemotherapeutic Drug List Available

Licensing Required For Commercial Use

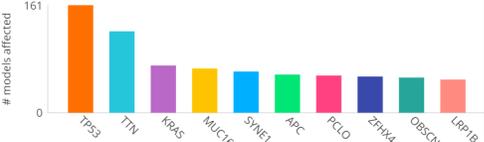
Race

← Use the filter panel on the left to customize your model search. SHARE VIEW LIST

Models By Primary Site
 26 Total

Has Multiple Models


2D Versus 3D Growth


Most Frequently Mutated Genes


Name	Primary Site	Clinical Tumor Diagnosis	Tissue Status	Age At Acquisition (Years)	Age At Diagnosis (Years)	Has Multiple Models	Expansion Status	# Mutated Genes	# Research Somatic Variants	# Clinical Variants	# Histo-Pathological Biomarkers
<input type="checkbox"/> HCM-BROD-0227-C43	Skin	Melanoma	Metastasis	40	40	No	EXPANDED	3075	4187	0	0
<input type="checkbox"/> HCM-SANG-0288-C18	Colon	Colorectal cancer	Primary	75		No	EXPANDED	3228	3908	0	0
<input type="checkbox"/> HCM-BROD-0569-C43	Skin	Melanoma	Metastasis	79	78	No	EXPANDED	2886	3802	0	1
<input type="checkbox"/> HCM-CSHL-0426-C18	Colon	Colorectal cancer	Primary	73	72	No	EXPANDED	2701	3183	0	0
<input type="checkbox"/> HCM-SANG-0273-C18	Colon	Colorectal cancer	Primary	78		No	EXPANDED	2597	2991	0	0
<input type="checkbox"/> HCM-BROD-0027-C34	Bronchus and lung	Lung cancer	Metastasis	66	65	No	EXPANDED	2313	2868	0	0
<input type="checkbox"/> HCM-CSHL-0459-C17	Small intestine	Rare cancers	Primary	57	57	No	EXPANDED	2426	2793	0	5
<input type="checkbox"/> HCM-BROD-0223-C43	Skin	Melanoma	Metastasis	74	73	No	EXPANDED	2187	2679	0	0
<input type="checkbox"/> HCM-SANG-0282-C18	Colon	Colorectal cancer	Primary	85		No	EXPANDED	2313	2636	0	0
<input type="checkbox"/> HCM-BROD-0106-C71	Brain	Glioblastoma	Recurrent	56	52	No	EXPANDED	2122	2333	0	3
<input type="checkbox"/> HCM-SANG-0276-C18	Colon	Colorectal cancer	Primary	78		No	EXPANDED	1768	1976	0	0
<input type="checkbox"/> HCM-BROD-0334-C43	Skin	Melanoma	Metastasis	72	70	No	EXPANDED	1619	1939	0	1
<input type="checkbox"/> HCM-CSHL-0174-C22	Intrahepatic bile duct	Intrahepatic bile duct cancer	Primary	64	64	No	EXPANDED	1568	1713	0	0
<input type="checkbox"/> HCM-BROD-0223-C43	Skin	Melanoma	Metastasis	70	69	No	EXPANDED	1367	1658	0	1
<input type="checkbox"/> HCM-CSHL-0317-C18	Colon	Colorectal cancer	Primary	65	64	No	EXPANDED	1502	1639	0	0
<input type="checkbox"/> HCM-CSHL-0729-C18	Colon	Colorectal cancer	Primary	60	60	No	EXPANDED	1433	1572	0	6
<input type="checkbox"/> HCM-BROD-0025-C16	Stomach	Stomach cancer	Primary	74	73	No	EXPANDED	1330	1422	0	0
<input type="checkbox"/> HCM-BROD-0681-C71	Brain	Glioblastoma	Recurrent	52	50	No	EXPANDED	1096	1166	0	3
<input type="checkbox"/> HCM-BROD-0679-C43	Skin	Melanoma	Metastasis	69	68	No	EXPANDED	765	878	0	1

Showing 1 - 20 of 295 models

Updated: July 31, 2023

« < 1 2 3 4 5 6 7 8 9 10 > »

Landing Page: Filter Panel and Interactive Search Results

Human Cancer Models Initiative
Searchable Catalog

Search By Model Name

Search By Altered Gene(s)

Primary Site

Pancreas

Colon

Rectum

Extrahepatic bile duct

Research Somatic Variant Type

Consequence

Model Type

3-D: Organoid

Has Multiple Models

Acquisition Site

Clinical Tumor Diagnosis

Clinical Stage Grouping

Histological Subtype

Histological Grade

Age At Diagnosis (Years)

Gender

Available Molecular Characterizations

Neoadjuvant Therapy

Chemotherapeutic Drug List Available

Licensing Required For Commercial Use

Race

CLEAR
MUTATED GENES is KRAS and PRIMARY SITE is Colon
SHARE
VIEW LIST

Models By Primary Site

Has Multiple Models

2D Versus 3D Growth

Most Frequently Mutated Genes

Name	Primary Site	Clinical Tumor Diagnosis	Tissue Status	Age At Acquisition (Years)	Age At Diagnosis (Years)	Has Multiple Models	Expansion Status	# Mutated Genes	# Research Somatic Variants	# Clinical Variants	# Histo-Pathological Biomarkers
HCM-SANG-0268-C18	Colon	Colorectal cancer	Primary	80		No	EXPANDED	223	233	0	0
HCM-SANG-0265-C18	Colon	Colorectal cancer	Metastasis	51		No	EXPANDED	208	219	0	0
HCM-CSHL-0247-C18-B	Colon	Colorectal cancer	Metastasis	76	76	Yes (2)	EXPANDED	209	216	1	5
HCM-SANG-0267-D12	Colon	Other	Primary	63		No	EXPANDED	196	203	0	0
HCM-CSHL-0056-C18	Colon	Colorectal cancer	Primary	75	75	No	EXPANDED	188	193	0	6
HCM-CSHL-0164-C20	Colon	Colorectal cancer	Metastasis	52	51	No	EXPANDED	171	177	0	0
HCM-SANG-0269-C18	Colon	Colorectal cancer	Primary	88		No	EXPANDED	158	165	0	0
HCM-CSHL-0247-C18-A	Colon	Colorectal cancer	Primary	76	76	Yes (2)	EXPANDED	160	162	1	5
HCM-CSHL-0063-C18	Colon	Colorectal cancer	Primary	75	75	No	EXPANDED	155	159	1	5
HCM-CSHL-0238-C18	Colon	Colorectal cancer	Primary	71	71	No	EXPANDED	154	156	0	5
HCM-SANG-0284-C18	Colon	Colorectal cancer	Primary	42		No	EXPANDED	141	143	0	0
HCM-CSHL-0461-D12	Colon	Other	Pre-malignant	54	54	No	EXPANDED	109	116	0	0
HCM-WCMC-0671-C18	Colon	Colorectal cancer	Primary	47	47	No	EXPANDED	110	111	1	7

Landing Page: Customized Data Visualization and Export

Human Cancer Models Initiative
Searchable Catalog

Search By Model Name

Search By Altered Gene(s)

Primary Site

Research Somatic Variant Type

Consequence

Model Type

Has Multiple Models

Acquisition Site

Clinical Tumor Diagnosis

Clinical Stage Grouping

Histological Subtype

Histological Grade

Age At Diagnosis (Years)

Gender

Available Molecular Characterizations

Neoadjuvant Therapy

Chemotherapeutic Drug List Available

Licensing Required For Commercial Use

Race

CLEAR MUTATED GENES is KRAS and PRIMARY SITE is Colon
SHARE VIEW LIST

Models By Primary Site

1 Total

Has Multiple Models

2D Versus 3D Growth

Most Frequently Mutated Genes

Showing 1 - 13 of 13 models

<input type="checkbox"/>	Name	Primary Site	Clinical Tumor Diagnosis	Tissue Status	Age At Acquisition (Years)	Age At Diag (Years)
<input type="checkbox"/>	HCM-SANG-0268-C18	Colon	Colorectal cancer	Primary	80	
<input type="checkbox"/>	HCM-SANG-0265-C18	Colon	Colorectal cancer	Metastasis	51	
<input type="checkbox"/>	HCM-CSHL-0247-C18-B	Colon	Colorectal cancer	Metastasis	76	76
<input type="checkbox"/>	HCM-SANG-0267-D12	Colon	Other	Primary	63	
<input type="checkbox"/>	HCM-CSHL-0056-C18	Colon	Colorectal cancer	Primary	75	75
<input type="checkbox"/>	HCM-CSHL-0164-C20	Colon	Colorectal cancer	Metastasis	52	51
<input type="checkbox"/>	HCM-SANG-0269-C18	Colon	Colorectal cancer	Primary	88	
<input type="checkbox"/>	HCM-CSHL-0247-C18-A	Colon	Colorectal cancer	Primary	76	76
<input type="checkbox"/>	HCM-CSHL-0063-C18	Colon	Colorectal cancer	Primary	75	75
<input type="checkbox"/>	HCM-CSHL-0238-C18	Colon	Colorectal cancer	Primary	71	71
<input type="checkbox"/>	HCM-SANG-0284-C18	Colon	Colorectal cancer	Primary	42	
<input type="checkbox"/>	HCM-CSHL-0461-D12	Colon	Other	Pre-malignant	54	54
<input type="checkbox"/>	HCM-WCMC-0671-C18	Colon	Colorectal cancer	Primary	47	47

Show 20 rows

COLUMNS

EXPORT

Histo-Pathological Biomarkers

0
5
0
6
0
0
5
5
5
0
0
7

✓ Select All ↺ Reset to Defaults

- Primary Site
- Clinical Tumor Diagnosis
- Histological Subtype
- Tissue Status
- Acquisition Site
- Gender
- Race
- Age At Acquisition (Years)
- Age At Diagnosis (Years)
- Disease Status
- Vital Status

Landing Page: Customized Data Visualization and Export

The screenshot displays the Human Cancer Models Initiative Searchable Catalog interface. On the left, a sidebar contains search filters for Model Name, Altered Gene(s), Research Somatic Variant, Primary Site, Research Somatic Variant Type, Consequence, Model Type, Has Multiple Models, Acquisition Site, Clinical Tumor Diagnosis, Clinical Stage Grouping, Histological Subtype, Histological Grade, Age At Diagnosis (Years), Gender, Available Molecular Characterizations, Neoadjuvant Therapy, Chemotherapeutic Drug List Available, Licensing Required For Commercial Use, and Race.

The main content area shows search results for models with MUTATED GENES is KRAS and PRIMARY SITE is Colon. A donut chart indicates 1 Total model. Below the chart is a table with columns: Name, Primary Site, Clinical Tumor Diagnosis, Tissue Status, and Age At Acquisition (Years). The table lists 13 models, with HCM-SANG-0267-D12, HCM-CSHL-0238-C18, and HCM-CSHL-0247-C18-A highlighted with red boxes.

On the right, the 'My Model List' panel shows three model cards, each with a thumbnail image and availability date. The 'EXPORT' button is highlighted with a red box. Below the list, a large yellow button labeled 'TSV (ALL COLUMNS)' is also highlighted with a red box. An orange arrow points to the 'VIEW LIST' button in the top right corner.

Name	Primary Site	Clinical Tumor Diagnosis	Tissue Status	Age At Acquisition (Years)
<input type="checkbox"/> HCM-SANG-0268-C18	Colon	Colorectal cancer	Primary	80
<input type="checkbox"/> HCM-SANG-0265-C18	Colon	Colorectal cancer	Metastasis	51
<input type="checkbox"/> HCM-CSHL-0247-C18-B	Colon	Colorectal cancer	Metastasis	76
<input checked="" type="checkbox"/> HCM-SANG-0267-D12	Colon	Other	Primary	63
<input type="checkbox"/> HCM-CSHL-0056-C18	Colon	Colorectal cancer	Primary	75
<input type="checkbox"/> HCM-CSHL-0164-C20	Colon	Colorectal cancer	Metastasis	52
<input type="checkbox"/> HCM-SANG-0269-C18	Colon	Colorectal cancer	Primary	88
<input checked="" type="checkbox"/> HCM-CSHL-0247-C18-A	Colon	Colorectal cancer	Primary	76
<input type="checkbox"/> HCM-CSHL-0063-C18	Colon	Colorectal cancer	Primary	75
<input checked="" type="checkbox"/> HCM-CSHL-0238-C18	Colon	Colorectal cancer	Primary	71
<input type="checkbox"/> HCM-SANG-0284-C18	Colon	Colorectal cancer	Primary	42
<input type="checkbox"/> HCM-CSHL-0461-D12	Colon	Other	Pre-malignant	54
<input type="checkbox"/> HCM-WCMC-0671-C18	Colon	Colorectal cancer	Primary	47

Landing Page: Navigating to Individual Model Pages

Human Cancer Models Initiative
Searchable Catalog

Search By Model Name

CLEAR
MUTATED GENES is KRAS and PRIMARY SITE is Colon

← SHARE
VIEW LIST

Models By Primary Site



1
Total

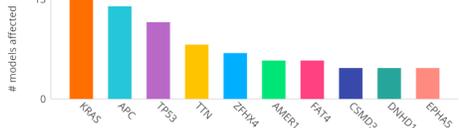
Has Multiple Models



2D Versus 3D Growth



Most Frequently Mutated Genes



Showing 1 - 13 of 13 models
Include 5 unexpanded models
COLUMNS
EXPORT

<input type="checkbox"/>	Name	Primary Site	Clinical Tumor Diagnosis	Tissue Status	Age At Acquisition (Years)	Age At Diagnosis (Years)	Has Multiple Models	Expansion Status	# Mutated Genes	# Research Somatic Variants	# Clinical Variants	# Histo-Pathological Biomarkers
<input type="checkbox"/>	HCM-SANG-0268-C18	Colon	Colorectal cancer	Primary	80		No	EXPANDED	223	233	0	0
<input type="checkbox"/>	HCM-SANG-0265-C18	Colon	Colorectal cancer	Metastasis	51		No	EXPANDED	208	219	0	0
<input type="checkbox"/>	HCM-CSHL-0247-C18-B	Colon	Colorectal cancer	Metastasis	76	76	Yes (2)	EXPANDED	209	216	1	5
<input checked="" type="checkbox"/>	HCM-SANG-0267-D12	Colon	Other	Primary	63		No	EXPANDED	196	203	0	0
<input type="checkbox"/>	HCM-CSHL-0056-C18	Colon	Colorectal cancer	Primary	75	75	No	EXPANDED	188	193	0	6
<input type="checkbox"/>	HCM-CSHL-0164-C20	Colon	Colorectal cancer	Metastasis	52	51	No	EXPANDED	171	177	0	0
<input type="checkbox"/>	HCM-SANG-0209-A18	Colon	Colorectal cancer	Primary	88		No	EXPANDED	158	165	0	0
<input checked="" type="checkbox"/>	HCM-CSHL-0247-C18-A	Colon	Colorectal cancer	Primary	76	76	Yes (2)	EXPANDED	160	162	1	5
<input type="checkbox"/>	HCM-CSHL-0069-C18	Colon	Colorectal cancer	Primary	75	75	No	EXPANDED	155	159	1	5
<input checked="" type="checkbox"/>	HCM-CSHL-0238-C18	Colon	Colorectal cancer	Primary	71	71	No	EXPANDED	154	156	0	5
<input type="checkbox"/>	HCM-SANG-0284-C18	Colon	Colorectal cancer	Primary	42		No	EXPANDED	141	143	0	0
<input type="checkbox"/>	HCM-CSHL-0461-D12	Colon	Other	Pre-malignant	54	54	No	EXPANDED	109	116	0	0
<input type="checkbox"/>	HCM-WCMC-0671-C18	Colon	Colorectal cancer	Primary	47	47	No	EXPANDED	110	111	1	7

Show 20 rows
Updated: July 25, 2023
« < 1 > »

Search By Altered Gene(s)

Search By Research Somatic Variant

Primary Site

- Pancreas 39
- Colon 13
- Rectum 7
- Extrahepatic bile duct 3
- 6 More

Research Somatic Variant Type

Consequence

Model Type

- 3-D: Organoid 13

Has Multiple Models

Acquisition Site

Clinical Tumor Diagnosis

Clinical Stage Grouping

Histological Subtype

Histological Grade

Age At Diagnosis (Years)

Gender

Available Molecular Characterizations

Neoadjuvant Therapy

Chemotherapeutic Drug List Available

Licensing Required For Commercial Use

Race

Individual Model Pages: Available Catalog Data for Selected Model

MODEL DETAILS

Model Type	3-D: Organoid
Split Ratio	N/A
Time to Split	N/A
Doubling Time	N/A
Tissue Status	Primary

MULTIPLE MODELS FROM THIS PATIENT (1)

HCM-CSHL-0247-C18-B
Tissue Status: Metastasis

AVAILABLE MOLECULAR CHARACTERIZATIONS (10)

	Model	Tumor	Normal
WGS	✓	✓	✓
WXS	✓	✓	✓
RNA-seq	✓	✓	✗
DNA Methylation	✓	✓	✗

PATIENT DETAILS

Tissue Status	Primary
Gender	Female
Race	White
Age At Diagnosis (Years)	76
Age At Acquisition (Years)	76
Disease Status	Stable disease
Vital Status	Alive
Neoadjuvant Therapy	No
Therapy	N/A
Chemotherapeutic Drug List Available	Yes
Clinical Tumor Diagnosis	Colorectal cancer
Histological Subtype	Adenocarcinoma
Primary Site	Colon
Acquisition Site	Cecum
TNM Stage	T3N2aM1a
Clinical Stage Grouping	Stage IVA
Histological Grade	G2

MODEL IMAGES (2)

©ATCC PDM-256™
Scale-bar length: 1000 µm | Magnification: 4 x

REPOSITORY STATUS

Date Updated	June 15, 2023
Date Of Availability	April 30, 2020
Licensing Required For Commercial Use	Yes
Date Created	May 08, 2020

EXTERNAL RESOURCES

📄 SEQUENCING FILE
📄 CASE METADATA
📄 MASKED SOMATIC MAF
🛒 VISIT PDM-256 TO PURCHASE

VARIANTS

Research Somatic Variants **Clinical Variants** are identified through clinical sequencing testing procedures as reported in the [case report forms](#).

Clinical Variants Showing 1 - 1 of 1 Variants Filter

Name	Genes	Type	Frequency
KRAS A146T	KRAS	SNV	2 <div style="width: 20px; height: 5px; background: linear-gradient(to right, #ccc, #ccc);"></div> 0.34%

External Resources: Available Data at GDC

CA HCMI-CMDC / HCM-CSHL-0247-C18

Add all files to the cart

Summary

Case UUID	7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8
Case ID	HCM-CSHL-0247-C18
Project	HCMI-CMDC
Project Name	NCI Cancer Model Development for the Human Cancer Model Initiative
Disease Type	Adenomas and Adenocarcinomas
Program	HCM I
Primary Site	Colon

FILES

132

ANNOTATIONS

0

File Counts by Data Category

Data Category	Files (n=132)
Sequencing Reads	22
Transcriptome Profiling	8
Simple Nucleotide Variation	64
Copy Number Variation	8
DNA Methylation	12
Clinical	1
Biospecimen	1

File Counts by Experimental Strategy

Experimental Strategy	Files (n=132)
WGS	29
WXS	61
RNA-Seq	28
Methylation Array	12

Clinical

Download Complete Set of Clinical Data

Demographic Diagnoses / Treatments (2) Family Histories (1) Exposures (1) Follow-Ups (15)

UUID	2055ab3d-7479-4645-b2b4-f4718cd4ac1d
Ethnicity	not hispanic or latino
Gender	female
Race	white
Days To Birth	-27936
Days To Death	--
Vital Status	Alive

Clinical Supplement File

Filename	Data format	Size	Action
HCM-CSHL-0247-C18.json	CDC JSON	169.89 KB	

External Resources: Available Data at GDC

CA HCMI-CMD / HCM-CSHL-0247-C18

[Download Complete Set of Clinical Data](#)

Demographic Diagnoses / Treatments (2) Family Histories (1) Exposures (1) Follow-Ups (15)



31bcc915	UUID	cde4e8e5-19c5-4576-9647-80d19ca2b15
	Classification Of Tumor	primary
	Alcohol Intensity	--
cde4e8e5...	Age At Diagnosis	76 years 177 days
	Days To Last Follow Up	--
	Days To Last Known Disease Status	--
	Days To Recurrence	--
	Last Known Disease Status	--
	Morphology	8140/3
	Primary Diagnosis	Adenocarcinoma, NOS
	Prior Malignancy	no
	Synchronous Malignancy	No
	Progression Or Recurrence	--
	Site Of Resection Of Biopsy	Cecum
	Tissue Or Organ Of Origin	Colon, NOS
	Tumor Grade	G2
	Tumor Stage	--

Treatments (12)

UUID	Therapeutic Agents	Treatment Intent Type	Treatment or Therapy	Days to Treatment Start
12b4c665-9161-43e7-8c8a-f9f02cad532e	--	Adjuvant	yes	155
3ceb476d-da26-4049-9e85-01383486cfae	--	Adjuvant	no	--
507740ed-a9c-42cb-a53e-50514dc1c882	--	--	yes	56
621b555f-451f-4d6c-81cf-5afda9709a28	Oxaliplatin	Adjuvant	yes	392
79f83b59-bf84-43e8-be08-6efaa5ab6e5c	Bevacizumab	Adjuvant	yes	155
8faa0b0f-541e-4673-94fb-ad1e45b036cd	--	Adjuvant	no	--
9ae03485-c2d9-4e15-a171-38d05aaa5e74	--	Neoadjuvant	no	--
b4e00b9e-2ded-4bce-8ab6-56190e790f04	Oxaliplatin	Adjuvant	yes	155
cafdb21c-4fc4-40b5-80dc-a1acd5cadb2	Leucovorin	Adjuvant	yes	392
df082093-a2b9-48a4-8814-486a8d36aec8	Leucovorin Calcium	Adjuvant	yes	155
f2f83b32-19a8-4529-b36c-3c32662f1391	Fluorouracil	Adjuvant	yes	155
fa3c33a2-da1e-4b17-b2d0-3191f946936a	Fluorouracil	Adjuvant	yes	392

Clinical Supplement File

Filename	Data format	Size	Action
HCM-CSHL-0247-C18.json	CDC JSON	169.89 KB	

External Resources: Available Data at GDC

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis Repository

Quick Search Manage Sets Login Cart GDC Apps

CA HCM-0247-C18 HCM-0247-C18.json CDC JSON 169.89 KB

Biospecimen

Download

Expand All

Sample ID	HCM-CSHL-0247-C18-10A
Sample UUID	72daf428-72a1-413c-8414-452beee8476f
Sample Type	Blood Derived Normal
Sample Type Id	10
Tissue Type	Normal
Tumor Code	--
Tumor Code Id	--
Oct Embedded	false
Shortest Dimension	--
Intermediate Dimension	--
Longest Dimension	--
Is Fpfe	false
Pathology Report Uuid	--
Tumor Descriptor	Not Applicable
Current Weight	--
Initial Weight	--
Composition	Buffy Coat
Time Between Clamping And Freezing	--
Time Between Excision And Freezing	--
Days To Sample Procurement	--
Freezing Method	--
Preservation Method	Frozen
Days To Collection	--
Portions	1

Biospecimen Supplement File

Filename	Data format	Size	Action
nationwidechildrens.org_biospecimen.HCM-CSHL-0247-C18.xml	BCR XML	80.38 KB	Download



Acknowledgments

- Center for Cancer Genomics
- Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc.
- NCI Contracting Office
- Center to Reduce Cancer Health Disparities

HCMI Consortium Founders:

- National Cancer Institute
- Wellcome Sanger Institute
- Cancer Research UK
- Hubrecht Organoid Technology

HCMI Searchable Catalog:

- Ontario Institute for Cancer Research
- Center for Biomedical Informatics and Information Technology

Cancer Model Development Centers:

- Broad Institute
- Cold Spring Harbor Laboratory
 - University of Verona and Hubrecht Institute
- Stanford University
- Weill Cornell Medical College

CMDC Model Development Pipeline:

- Biospecimen Processing Center: Nationwide Children's Hospital
- Clinical Data Center: Information Management Systems
- NCI Data Coordinating Center
- Genome Characterization Centers: Broad Institute and University of North Carolina
- Genomic Data Commons (GDC)

HCMI Model Distributor:

- ATCC



**NATIONAL
CANCER
INSTITUTE**

www.cancer.gov

www.cancer.gov/espanol

HCMI Data in the Genomic Data Commons

17 August 2023

Bill Wysocki, Ph.D. – GDC User Services Lead
Center for Translational Data Science
University of Chicago

HCMI Data in the GDC

- *Brief Introduction to the GDC*
- *Available HCMI Data*
- *HCMI Data Download Walkthrough*
- *Controlled Data Access*

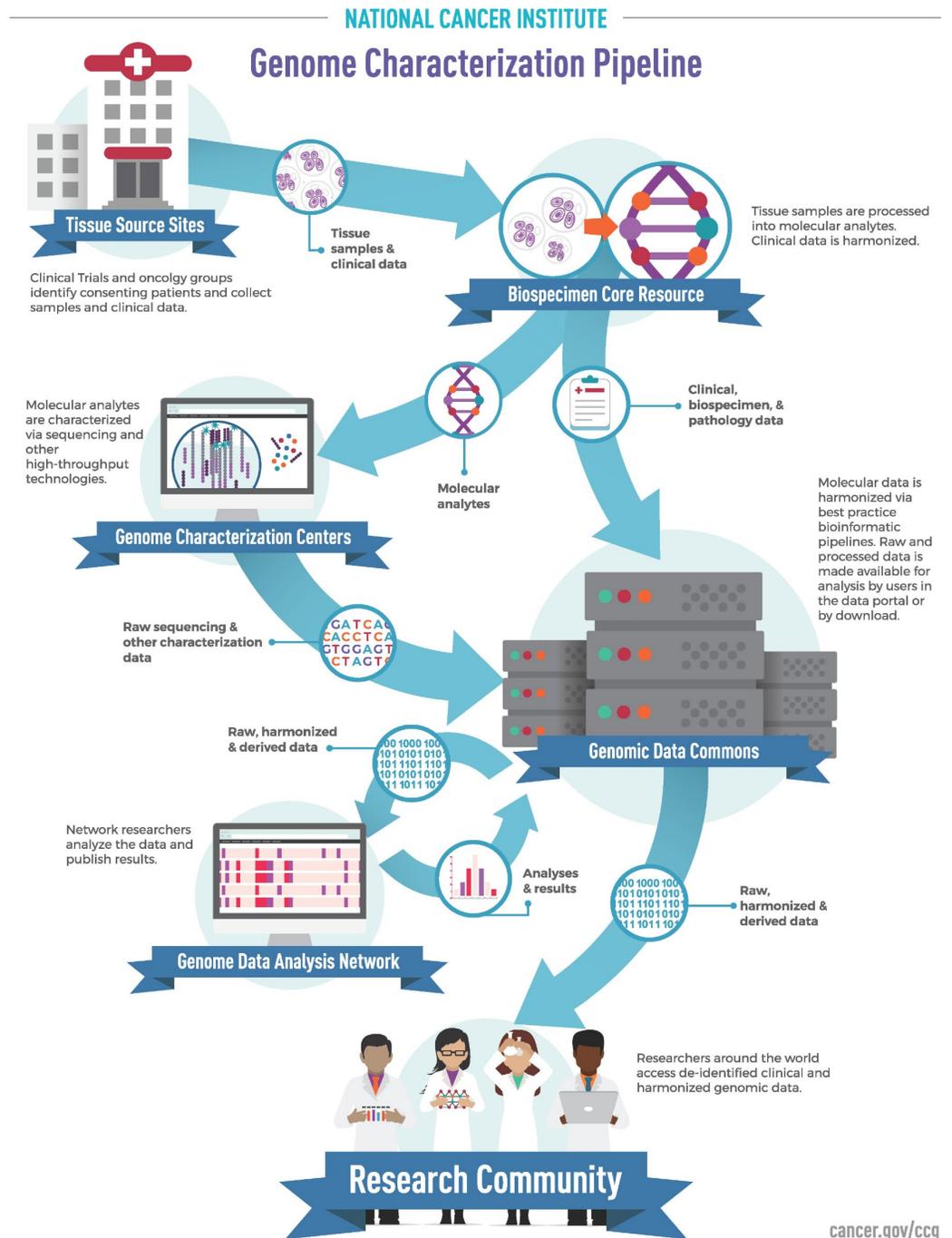
Introduction to the GDC

The Genomic Data Commons

The GDC is a data repository for cancer genomics. This includes molecular, biospecimen, and clinical data.

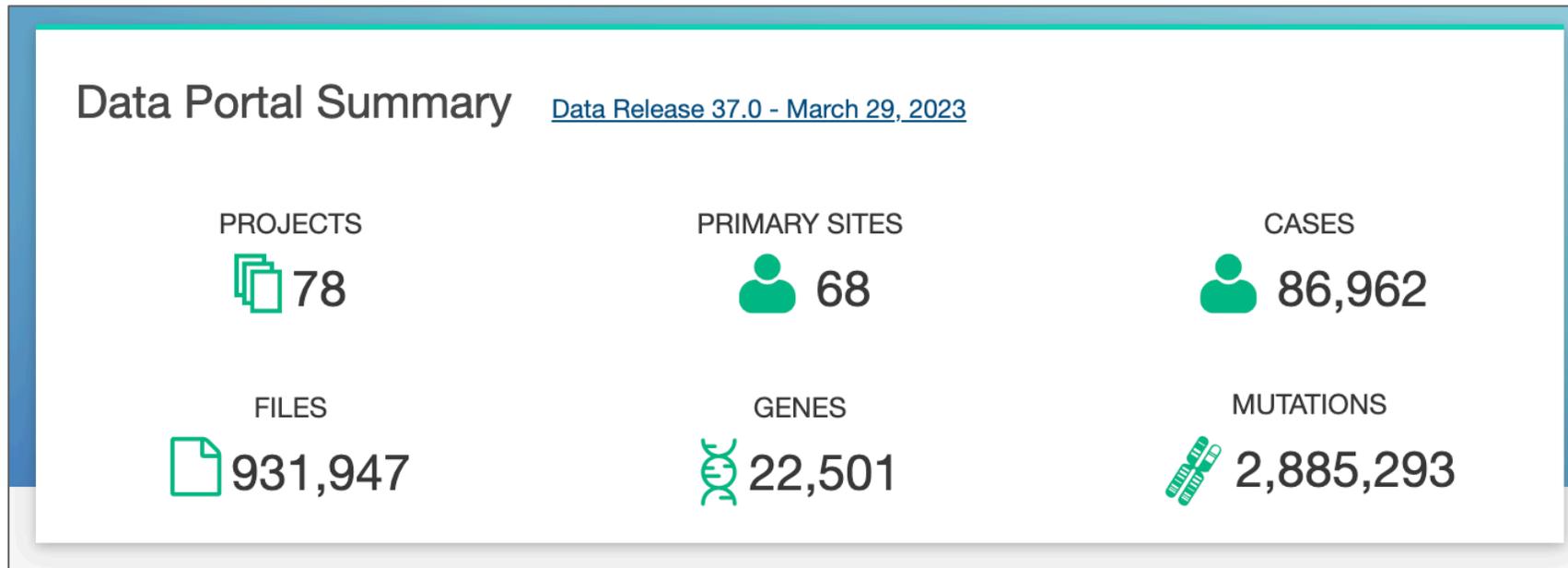
Data is processed using a set of standardized pipelines and stored in a flexible data model.

Data is submitted by researchers to be processed and eventually released to the wider research community



GDC Data Summary – August 2023 – Data Release 37

~3.8 PB of data across **932K data files**
86,962 cases across **78 projects**
2.9 million explorable somatic mutations



GDC Data

Molecular data – Includes data from multiple experimental strategies. Harmonized data starts at alignment and includes downstream derived data.

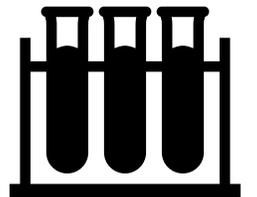
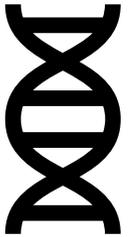
WXS, WGS, RNA-Seq, Methylation

Clinical data – Uploaded as metadata associated with cases (patients) but is also available as supplement files.

Demographic, Diagnosis, Exposure, Family History, Pathology

Biospecimen data – Also uploaded as metadata, but is associated with samples, portions, analytes, and aliquots. Also available as supplement files.

Sample, Portion, Analyte, Aliquot, Slides

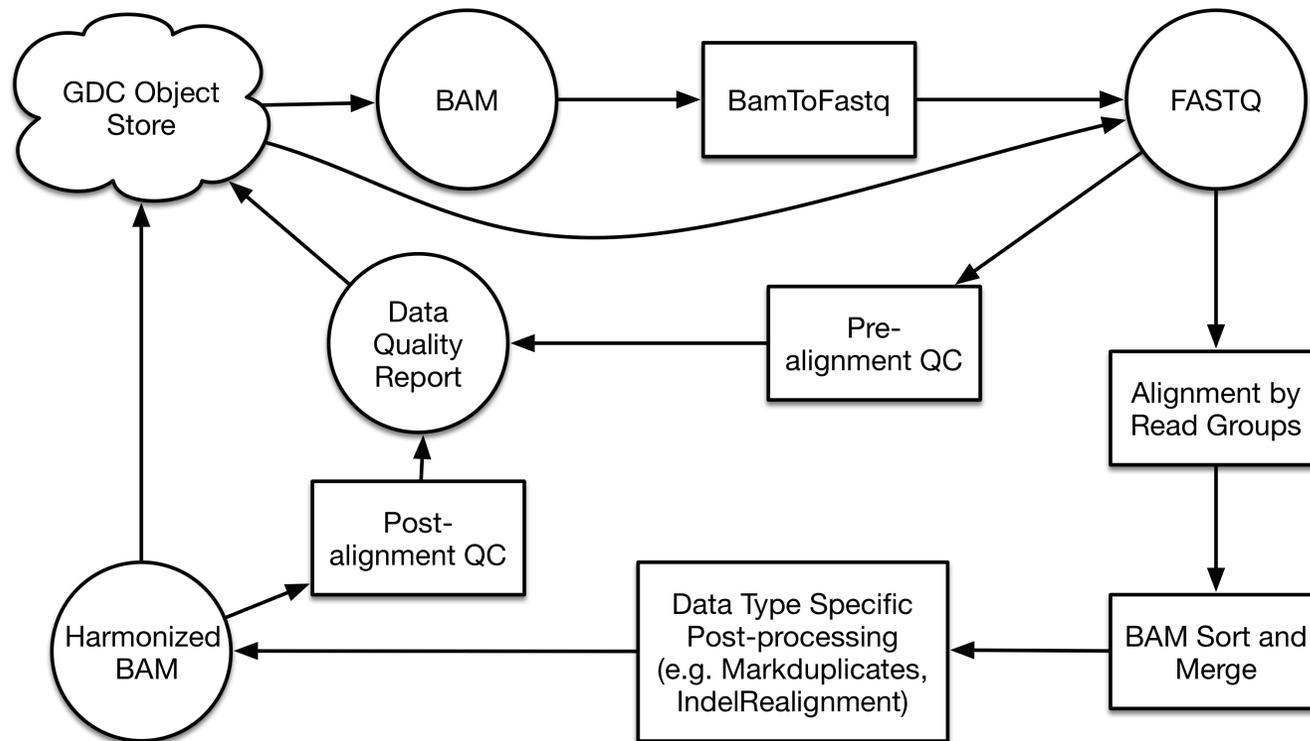


HCMI Data Available in the GDC

Harmonized Data

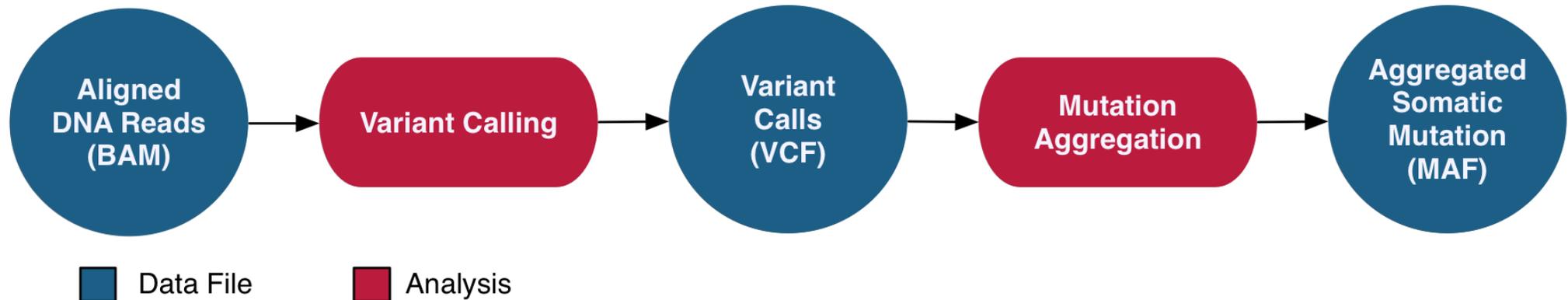
Molecular Data Strategies Available for HCMI

- Alignments available for all NGS strategies
 - All samples aligned to the same reference genome



Molecular Data Strategies Available for HCMI

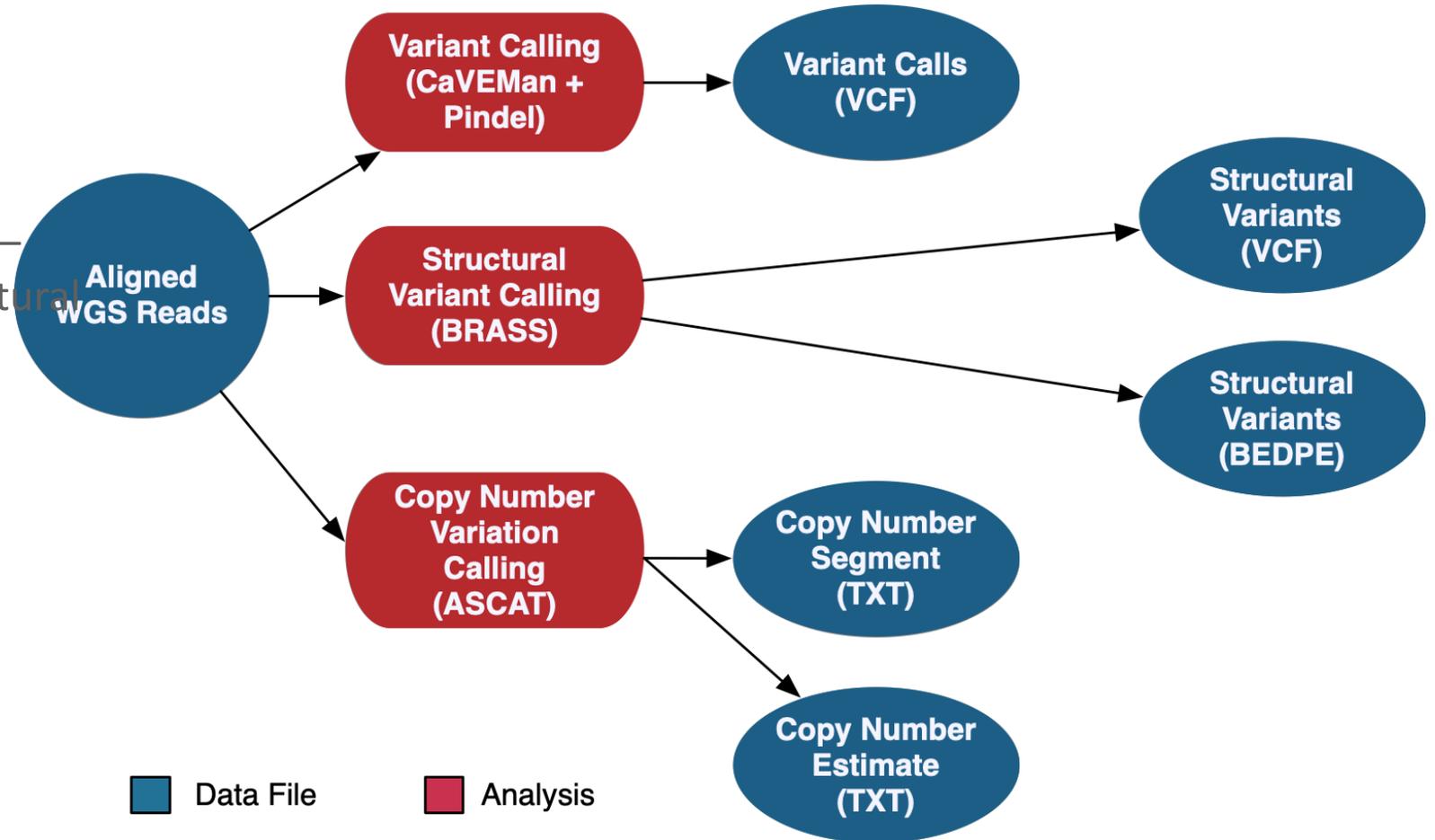
- **Whole Exome Sequencing**
 - Somatic variant calling pipeline – SNVs and indels
 - Available raw or filtered



Molecular Data Strategies Available for HCMI

- **Whole Genome Sequencing**

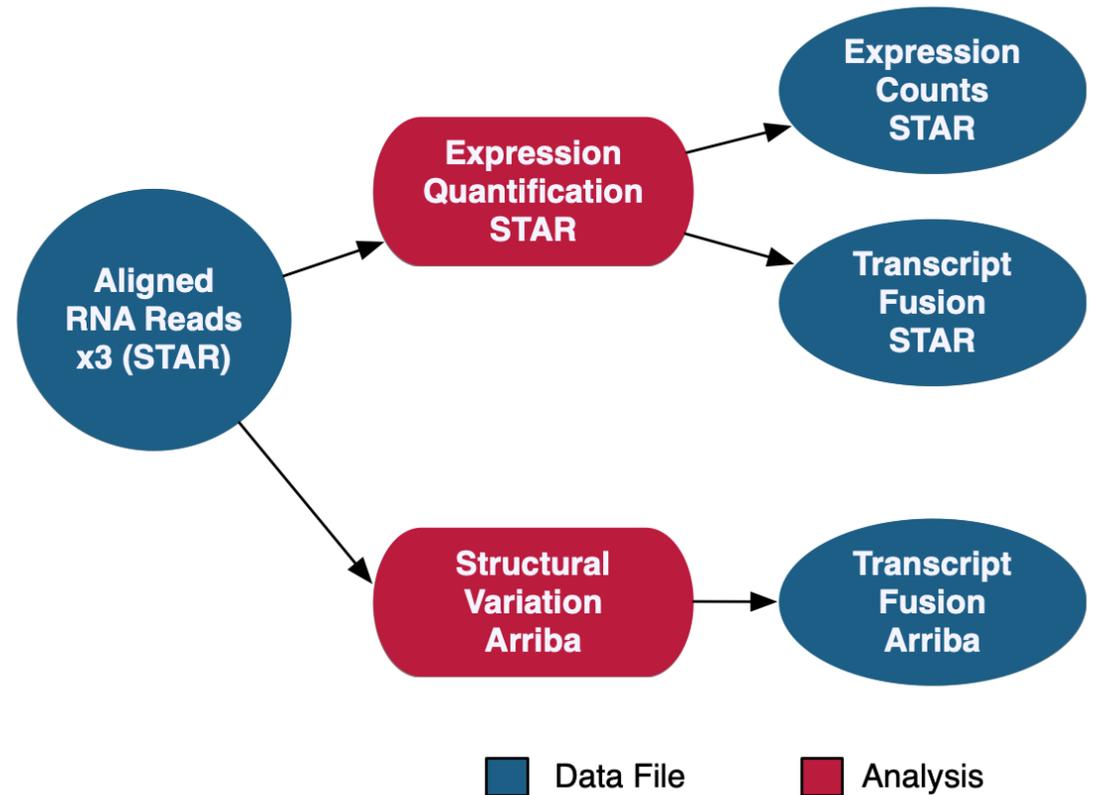
- Somatic variant calling pipelines – SNVs, indels, copy number, structural



Molecular Data Strategies Available for HCM1

- **RNA-Seq**

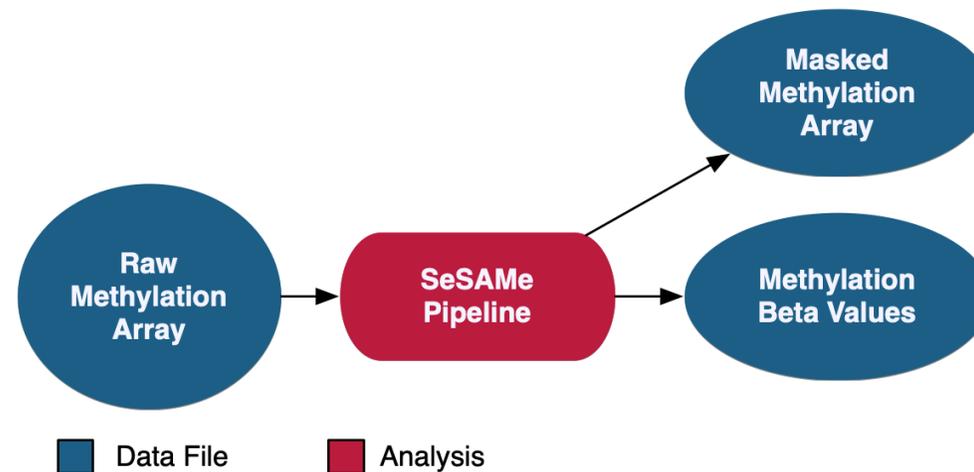
- Gene expression values available – Read counts per gene and normalized.
- Transcript Fusion and Splice Junction files are also available.



Molecular Data Strategies Available for HCMII

- **Methylation Array**

- Raw array files available with germline probes removed.
- Methylation Beta Values are also available.



HCMI Data Download Walkthrough

Use Case: Downloading RNA-Seq Gene Expression Data

HCMCI Case Page

CA HCMCI-CMDC / HCM-CSHL-0247-C18

 Add all files to the cart

Summary

Case UUID	7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8
Case ID	HCM-CSHL-0247-C18
Project	HCMCI-CMDC
Project Name	NCI Cancer Model Development for the Human Cancer Model Initiative
Disease Type	Adenomas and Adenocarcinomas
Program	HCMCI
Primary Site	Colon

FILES
132



ANNOTATIONS
0



File Counts by Data Category

Data Category	Files (n=132)
Sequencing Reads	22
Transcriptome Profiling	8
Simple Nucleotide Variation	64
Copy Number Variation	8
DNA Methylation	12
Clinical	1
Biospecimen	1

File Counts by Experimental Strategy

Experimental Strategy	Files (n=132)
WGS	29
WXS	61
RNA-Seq	28
Methylation Array	12

HCMCI Case Page

CA HCMI-CMDC / HCM-CSHL-0247-C18

Add all files to the cart

Summary

Case UUID	7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8
Case ID	HCM-CSHL-0247-C18
Project	HCMI-CMDC
Project Name	NCI Cancer Model Development for the Human Cancer Model Initiative
Disease Type	Adenomas and Adenocarcinomas
Program	HCMCI
Primary Site	Colon

RNA-Seq

28

FILES
132

ANNOTATIONS
0

File Counts by Data Category

Data Category	Files (n=132)
Sequencing Reads	22
Transcriptome Profiling	8
Simple Nucleotide Variation	64
Copy Number Variation	8
DNA Methylation	12
Clinical	1
Biospecimen	1

File Counts by Experimental Strategy

Experimental Strategy	Files (n=132)
WGS	29
WXS	61
RNA-Seq	28
Methylation Array	12

HCMCI Case Page

CA HCMCI-CMDC / HCM-CSHL-0247-C18

Add all files to the cart

Summary

Case UUID	7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8
Case ID	HCM-CSHL-0247-C18
Project	HCMCI-CMDC
Project Name	NCI Cancer Model Development for the Human Cancer Model Initiative
Disease Type	Adenomas and Adenocarcinomas
Program	HCMCI
Primary Site	Colon

RNA-Seq

28

FILES
132

ANNOTATIONS
0

File Counts by Data Category

Data Category	Files (n=132)
Sequencing Reads	22
Transcriptome Profiling	8
Simple Nucleotide Variation	64
Copy Number Variation	8
DNA Methylation	12
Clinical	1
Biospecimen	1

File Counts by Experimental Strategy

Experimental Strategy	Files (n=132)
WGS	29
WXS	61
RNA-Seq	28
Methylation Array	12

HCMCI Biospecimen Check

Biospecimen Download

Q RNA Collapse All

Analyte ID	HCM-CSHL-0247-C18-85A-01R
Analyte UUID	daf1194b-6af2-492d-b074-342150aee012
Analyte Type	RNA
Analyte Type Id	R
Well Number	--
Amount	--
A260 A280 Ratio	1.97
Concentration	0.15
Spectrophotometer Method	UV Spec
Aliquots	1

1 HCM-CSHL-0247-C18-85A-01R

2 HCM-CSHL-0247-C18-01A-21R

3 HCM-CSHL-0247-C18-85M-01R

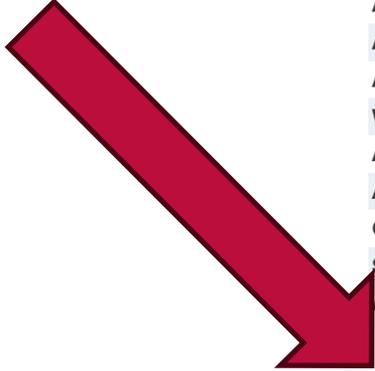
4 RNA-Seq Samples

Biospecimen

Q RNA x

[-] Samples

- [-] HCM-CSHL-0247-C18-10A
 - [+] Portions



Repository View: Files associated with Case

Files
Cases
«
[Browse Annotations](#)

Clear
Case
IS 7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8
AND
Experimental Strategy
IS
RNA-Seq
Advanced Search

Files (28)
Cases (1)
Add All Files to Cart
Manifest
View 1 Case in Exploration
View Images

Primary Site

Project

Data Category

Data Type

Data Format

Show More

Showing 1 - 20 of 28 files
131.43 GB

☰
↓
JSON
TSV

File UUID	File Name	Cases	Project	Access	Data Category	Data Format	File Size	Annotations
24208746-40da-4563-b108-b4cd753934	45939b89-9b0f-483f-bce0-e1a150778d5c.rna_seq.genomic.gdc_realn.bam	1	HCMI-CMDC	🔒 controlled	Sequencing Reads	BAM	16.89 GB	0
2c98666e-64af-4eb8-b6e6-2da66720a7be	HCMI-CMDC.7c6c769a-ffc2-46bb-bb62-50f7026293f1.star_fusion.rna_fusion.bedpe	1	HCMI-CMDC	🔒 controlled	Structural Variation	BEDPE	661 B	0

Search Files
?

🔍 e.g. 142682.bam, 4f6e2e7a-b...

> Data Category
↻

> Data Type

> Experimental Strategy

WXS
 WGS
 RNA-Seq
 Methylation Array

Files
61
29
28
12

> Workflow Type

> Data Format

> Platform

> Access

Repository View: Files associated with Case

Files
Cases
«
[Browse Annotations](#)

[Add a File Filter](#)

Search Files ?

> Data Category

> Data Type

> Experimental Strategy ↻

	# Files
<input type="checkbox"/> WXS	61
<input type="checkbox"/> WGS	29
<input checked="" type="checkbox"/> RNA-Seq	28
<input type="checkbox"/> Methylation Array	12

> Workflow Type

> Data Format

> Platform

> Access

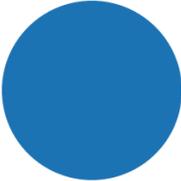
← Clear
Case
IS
7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8
AND
Experimental Strategy
IS
RNA-Seq
⚙️ Advanced Search

Files (28)

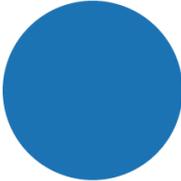
Cases (1)

Add All Files to Cart
Manifest
View 1 Case in Exploration
View Images

Primary Site



Project



Data Category



Data Type



Data Format



[Show More](#)

Showing 1 - 20 of 28 files 📄 131.43 GB

☰
⌵
JSON
TSV

	File UUID	File Name	Cases	Project	Access	Data Category	Data Format	File Size	Annotations
	24208746-40da-456	45939b89-9b0f-483f-bce0-e1a150							
	3-b108-b4cd753934	778d5c.rna_seq.genomic.gdc_real 4d n.bam	1	HCMI-CMDC	🔒 controlled	Sequencing Reads	BAM	16.89 GB	0
	2c98666e-64af-4eb8	HCMI-CMDC.7c6c769a-fbc2-46bb -b6e6-2da66720a7b -bb62-50f7026293f1.star_fusion.rn e a_fusion.bedpe	1	HCMI-CMDC	🔒 controlled	Structural Variation	BEDPE	661 B	0

Repository View: More Filters

Files Cases « 🔗 Browse Annotations

[Add a File Filter](#)

Search Files ?

🔍 e.g. 142682.bam, 4f6e2e7a-b...

> Data Category

▼ Data Type

- Aligned Reads # Files 12
- Transcript Fusion 8
- Gene Expression Quantification 4
- Splice Junction Quantification 4

> Experimental Strategy

▼ Workflow Type

- STAR - Counts 8
- Arriba 4
- STAR 2-Pass Chimeric 4
- STAR 2-Pass Genome 4

Clear Case IS 7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8 AND Experimental Strategy IS RNA-Seq ⚙️ Advanced Search

Files (28) Cases (1) 🛒 Add All Files to Cart 📄 Manifest 🔍 View 1 Case in Exploration 🖼 View Images

Primary Site Project Data Category Data Type Data Format

Show More

Showing 1 - 20 of 28 files 📁 131.43 GB ☰ ⚙️ JSON TSV

File UUID	File Name	Cases	Project	Access	Data Category	Data Format	File Size	Annotations
24208746-40da-456	45939b89-9b0f-483f-bce0-e1a150							
3-b108-b4cd753934	778d5c.rna_seq.genomic.gdc_real 4d n.bam	1	HCMI-CMDC	🔒 controlled	Sequencing Reads	BAM	16.89 GB	0
2c98666e-64af-4eb8	HCMI-CMDC.7c6c769a-fbc2-46bb							
-b6e6-2da66720a7b	-bb62-50f7026293f1.star_fusion.rn a_fusion.bedpe	1	HCMI-CMDC	🔒 controlled	Structural Variation	BEDPE	661 B	0

Repository View: More Filters

Files Cases << [Browse Annotations](#)

[Add a File Filter](#)

Search Files ?
e.g. 142682.bam, 4f6e2e7a-b...

> Data Category

> Data Type 

- Aligned Reads # Files 12
- Transcript Fusion # Files 8
- Gene Expression Quantification # Files 4
- Splice Junction Quantification # Files 4

> Experimental Strategy 

> Workflow Type

- STAR - Counts # Files 4

> Data Format

> Platform

> Access

Clear Case IS 7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8 AND

Data Type IS Gene Expression Quantification AND Experimental Strategy IS RNA-Seq [Advanced Search](#)

Files (4) Cases (1) [Add All Files to Cart](#) [Manifest](#) [View 1 Case in Exploration](#) [View Images](#)

Primary Site Project Data Category Data Type Data Format

[Show More](#)

Showing 1 - 4 of 4 files  16.89 MB   JSON TSV

 File UUID	File Name	Cases	Project	Access	Data Category	Data Format	File Size	Annotations
 77528506-9698-483e-b019-d8b005aa3326	f900d2a4-f3a6-4268-97ab-319307ff37d6.rna_seq.augmented_star_gene_counts.tsv	1	HCMI-CMD	 open	Transcriptome Profiling	TSV	4.22 MB	0
 0228ad1c-d5ac-4938-0157-4f7300-1f000	1016d0b2-d30e-4c8d-bcb5-1a5f0b991f2e.rna_seq.augmented_star_gene_c	1	HCMI-CMD	 open	Transcriptome Profiling	TSV	4.23 MB	0

Repository View: More Filters

Files Cases << [Browse Annotations](#)

[Add a File Filter](#)

Search Files ?
e.g. 142682.bam, 4f6e2e7a-b...

> Data Category

< Data Type 

- Aligned Reads # Files 12
- Transcript Fusion # Files 8
- Gene Expression Quantification # Files 4
- Splice Junction Quantification # Files 4

> Experimental Strategy 

< Workflow Type

- STAR - Counts # Files 4

> Data Format

> Platform

> Access

Clear Case IS 7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8 AND

Data Type IS Gene Expression Quantification AND Experimental Strategy IS RNA-Seq [Advanced Search](#)

Files (4) Cases (1) [Add All Files to Cart](#) [Manifest](#) [View 1 Case in Exploration](#) [View Images](#)

Primary Site Project Data Category Data Type Data Format

[Show More](#)

Showing 1 - 4 of 4 files  16.89 MB   JSON TSV

 File UUID	File Name	Cases	Project	Access	Data Category	Data Format	File Size	Annotations
 77528506-9698-483e-b019-d8b005aa3326	f900d2a4-f3a6-4268-97ab-319307ff37d6.rna_seq.augmented_star_gene_counts.tsv	1	HCMI-CMD	 open	Transcriptome Profiling	TSV	4.22 MB	0
 0228ad1c-d5ac-4938-0157-4f7300-1f000	1016d0b2-d30e-4c8d-bcb5-1a5f0b991f2e.rna_seq.augmented_star_gene_c	1	HCMI-CMD	 open	Transcriptome Profiling	TSV	4.23 MB	0

Repository View: Add Files to Cart

The screenshot displays the NIH GDC Data Portal interface. At the top, the NIH logo and 'NATIONAL CANCER INSTITUTE GDC Data Portal' are visible. Navigation links include Home, Projects, Exploration, Analysis, and Repository (which is highlighted). Utility links for Quick Search, Manage Sets, Login, Cart (0), and GDC Apps are also present.

The main content area is divided into a left sidebar and a main panel. The sidebar contains tabs for 'Files' and 'Cases', and a search bar with the text 'e.g. 142682.bam, 4f6e2e7a-b...'. Below the search bar are filter sections for 'Data Category', 'Data Type', 'Experimental Strategy', and 'Workflow Type'. The 'Data Type' section is expanded, showing checkboxes for 'Aligned Reads' (12 files), 'Transcript Fusion' (8 files), 'Gene Expression Quantification' (4 files, checked), and 'Splice Junction Quantification' (4 files). The 'Workflow Type' section shows 'STAR - Counts' (4 files).

The main panel features a search bar with filters: 'Case IS 7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8 AND Data Type IS Gene Expression Quantification AND Experimental Strategy IS RNA-Seq'. An 'Advanced Search' button is located to the right. Below the search bar, there are buttons for 'Add All Files to Cart' (highlighted with a red box), 'Manifest', 'View 1 Case in Exploration', and 'View Images'. A grid of five blue circles represents file categories: Primary Site, Project, Data Category, Data Type, and Data Format. A 'Show More' button is positioned below the 'Data Category' circle.

At the bottom, a summary bar indicates 'Showing 1 - 4 of 4 files' and '16.89 MB'. A table header is visible with columns: File UUID, File Name, Cases Project, Access Data Category, Data Format, File Size, and Annotations. The table content is partially obscured by a dashed line.

Repository View: Add Files to Cart

The screenshot displays the NIH GDC Data Portal interface. At the top, the navigation bar includes the NIH logo, 'NATIONAL CANCER INSTITUTE GDC Data Portal', and menu items for Home, Projects, Exploration, Analysis, and Repository. The 'Repository' menu is highlighted. On the right, there are links for Quick Search, Manage Sets, Login, and a 'Cart' icon with a '4' inside, which is highlighted with a red box. A 'GDC Apps' menu is also visible.

The main content area is divided into a left sidebar and a main panel. The sidebar has tabs for 'Files' and 'Cases'. It contains a search bar with the text 'e.g. 142682.bam, 4f6e2e7a-b...' and several filter sections: 'Data Category', 'Data Type' (with checkboxes for Aligned Reads (12), Transcript Fusion (8), Gene Expression Quantification (4), and Splice Junction Quantification (4)), 'Experimental Strategy', and 'Workflow Type' (with a checkbox for STAR - Counts (4)).

The main panel features a search bar with filters: 'Case IS 7fa8bc57-4bc7-48a5-99fe-98ed02fa17e8 AND Data Type IS Gene Expression Quantification AND Experimental Strategy IS RNA-Seq'. There is an 'Advanced Search' button and a 'Browse Annotations' link. Below the search bar, there are buttons for 'Add All Files to Cart', 'Manifest', 'View 1 Case in Exploration', and 'View Images'. A summary row shows 'Files (4)' and 'Cases (1)'. Below this, five blue circles represent filters for 'Primary Site', 'Project', 'Data Category', 'Data Type', and 'Data Format'. A 'Show More' button is located below the 'Data Category' circle. At the bottom, it says 'Showing 1 - 4 of 4 files' and '16.89 MB'. There are icons for menu, sort, and export (JSON, TSV). A table header is visible at the bottom with columns: File UUID, File Name, Cases Project, Access Data Category, Data Format, File Size, and Annotations.

Cart: Download Files

NATIONAL CANCER INSTITUTE
GDC Data Portal

Home
Projects
Exploration
Analysis
Repository

Manage Sets
4

FILES
4

CASES
1

FILE SIZE
16.89 MB

File Counts by Project

Project	Cases (n=1)	Files (n=4)	File Size (Σ=16.89 MB)
HCMI-CMDC	1	4	16.89 MB

File Counts by Authorization Level

Level	Files (n=4)	File Size (Σ=16.89 MB)
Authorized	4	16.89 MB

How to download files in my Cart?

Download Manifest:
Download a manifest for use with the [GDC Data Transfer Tool](#). The GDC Data Transfer Tool is recommended for transferring large volumes of data.

Download Cart:
Download Files in your Cart directly from the Web Browser.

Download Reference Files:
Download [GDC Reference Files](#) for use in your genomic data analysis.

Biospecimen
Clinical
Sample Sheet
Metadata
Download ▾
Remove From Cart ▾

Cart Items

Showing 1 - 4 of 4 files

Remove	File UUID	File Name	Cases	Project	Access	Data Category	Data Format	File Size	Annotations
	77528506-9698-483e-b019-d8b005aa3326	f900d2a4-f3a6-4268-97ab-319307ff37d6.rna_seq.augmented_star_gene_counts.tsv	1	HCMI-CMDC	open	Transcriptome Profiling	TSV	4.22 MB	0
	0228ad1c-d5ac-4938-9157-4f7738edf233	1016d0b2-d30e-4c8d-bcb5-1a5f0b991f2e.rna_seq.augmented_star_gene_counts.tsv	1	HCMI-CMDC	open	Transcriptome Profiling	TSV	4.23 MB	0

NIH NATIONAL CANCER INSTITUTE

46

Cart: Download Files

NATIONAL CANCER INSTITUTE
GDC Data Portal

[Home](#)
[Projects](#)
[Exploration](#)
[Analysis](#)
[Repository](#)

[Manage Sets](#)
🛒 4

FILES
4

CASES
1

FILE SIZE
16.89 MB

File Counts by Project

Project	Cases (n=1)	Files (n=4)	File Size (Σ=16.89 MB)
HCMI-CMDC	1	4	16.89 MB

File Counts by Authorization Level

Level	Files (n=4)	File Size (Σ=16.89 MB)
Authorized	4	16.89 MB

How to download files in my Cart?

Download Manifest:
Download a manifest for use with the [GDC Data Transfer Tool](#). The GDC Data Transfer Tool is recommended for transferring large volumes of data.

Download Cart:
Download Files in your Cart directly from the Web Browser.

Download Reference Files:
Download [GDC Reference Files](#) for use in your genomic data analysis.

📄 Biospecimen
📄 Clinical
📄 Sample Sheet
📄 Metadata
📄 Download ▾
🗑️ Remove From Cart ▾

Cart Items

Showing 1 - 4 of 4 files

Remove	File UUID	File Name	Cases	Project	Access	Data Category	Data Format	File Size	Annotations
	77528506-9698-483e-b019-d8b005aa3326	f900d2a4-f3a6-4268-97ab-319307ff37d6.rna_seq.augmented_star_gene_counts.tsv	1	HCMI-CMDC	🔒 open	Transcriptome Profiling	TSV	4.22 MB	0
	0228ad1c-d5ac-4938-9157-4f7738edf233	1016d0b2-d30e-4c8d-bcb5-1a5f0b991f2e.rna_seq.augmented_star_gene_counts.tsv	1	HCMI-CMDC	🔒 open	Transcriptome Profiling	TSV	4.23 MB	0

Cart: Download Files

NIH NATIONAL CANCER INSTITUTE GDC Data Portal [Home](#) [Projects](#) [Exploration](#) [Analysis](#) [Repository](#) [Manage Sets](#) [4](#)

FILES
4

CASES
1

FILE SIZE
16.89 MB

File Counts by Project

Project	Cases (n=1)	Files (n=4)	File Size (Σ=16.89 MB)
HCMI-CMDC	1	4	16.89 MB

File Counts by Authorization Level

Level	Files (n=4)	File Size (Σ=16.89 MB)
Authorized	4	16.89 MB

How to download files in my Cart?

Download Manifest:
Download a manifest for use with the [GDC Data Transfer Tool](#). The GDC Data Transfer Tool is recommended for transferring large volumes of data.

Download Cart:
Download Files in your Cart directly from the Web Browser.

Download Reference Files:
Download [GDC Reference Files](#) for use in your genomic data analysis.

[Biospecimen](#) [Clinical](#) [Sample Sheet](#) [Metadata](#) [Download](#) [Remove From Cart](#)

Cart Items

Showing 1 - 4 of 4 files

[Manifest](#)
[Cart](#)

[≡](#) [↓](#) [TSV](#)

Cart: Download File Metadata – Clinical & Biospecimen

 **NATIONAL CANCER INSTITUTE**
GDC Data Portal

[Home](#) [Projects](#) [Exploration](#) [Analysis](#) [Repository](#)

[Manage Sets](#)  

FILES
4

CASES
1

FILE SIZE
16.89 MB

File Counts by Project

Project	Cases (n=1)	Files (n=4)	File Size (Σ=16.89 MB)
HCMC-CMDC	1 	4 	16.89 MB 

File Counts by Authorization Level

Level	Files (n=4)	File Size (Σ=16.89 MB)
Authorized	4 	16.89 MB 

How to download files in my Cart?

Download Manifest:
Download a manifest for use with the [GDC Data Transfer Tool](#). The GDC Data Transfer Tool is recommended for transferring large volumes of data.

Download Cart:
Download Files in your Cart directly from the Web Browser.

Download Reference Files:
Download [GDC Reference Files](#) for use in your genomic data analysis.

Cart Items

Showing 1 - 4 of 4 files

Cart: Download File Metadata – Other

NIH NATIONAL CANCER INSTITUTE GDC Data Portal Home Projects Exploration Analysis Repository Manage Sets 4

FILES 4
CASES 1
FILE SIZE 16.89 MB

File Counts by Project

Project	Cases (n=1)	Files (n=4)	File Size (Σ=16.89 MB)
HCMC-CMDC	1	4	16.89 MB

File Counts by Authorization Level

Level	Files (n=4)	File Size (Σ=16.89 MB)
Authorized	4	16.89 MB

How to download files in my Cart?

Download Manifest:
Download a manifest for use with the [GDC Data Transfer Tool](#). The GDC Data Transfer Tool is recommended for transferring large volumes of data.

Download Cart:
Download Files in your Cart directly from the Web Browser.

Download Reference Files:
Download [GDC Reference Files](#) for use in your genomic data analysis.

Cart Items
Showing 1 - 4 of 4 files

[Biospecimen](#) [Clinical](#) [Sample Sheet](#) [Metadata](#) [Download](#) [Remove From Cart](#)

[TSV](#)
[JSON](#)

Basic Metadata (TSV) **All Possible Metadata (JSON)**

≡ ↓ TSV

Opening Downloaded Files

File will download as a TAR.GZ file

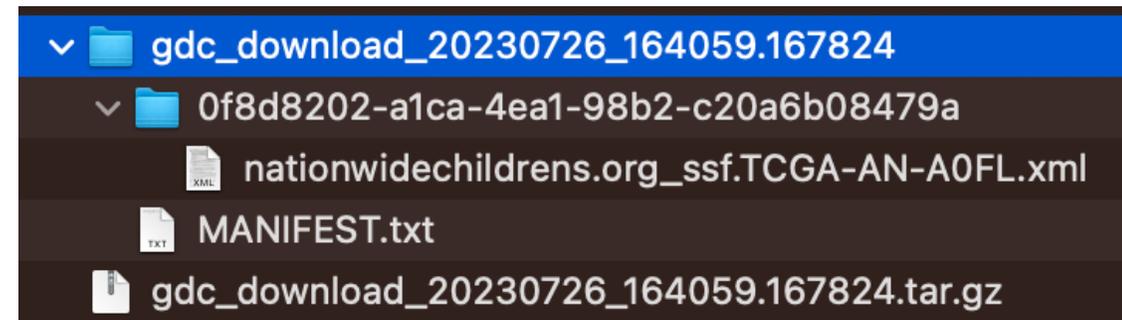
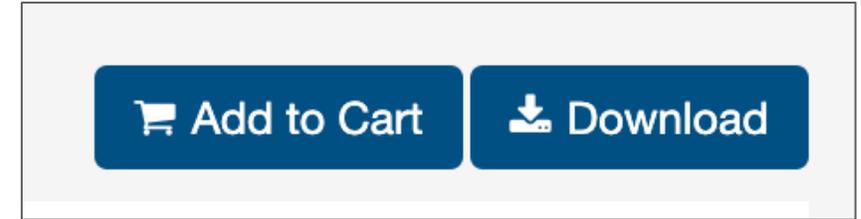
MAC OS will open these files automatically upon download

Windows will require another program

Two most common are **WinRAR** and **7-Zip**

When cart is downloaded, each file will be in a folder named after its UUID.

Will also contain a manifest and any annotations associate with files.



Two Additional Download Methods

The screenshot shows the GDC Data Portal interface. A red box highlights the 'Access' section in the left sidebar, which contains two radio button options: 'controlled' and 'open'. Below the sidebar, a table lists files with columns for 'File UUID', 'File Name', and 'Cases Project'. A red arrow points from the 'Access' section to a larger view of the same section at the bottom of the slide.

File UUID	File Name	Cases Project
24208746-40d	45939b89-9b0f-483f-	

Data Portal

The diagram illustrates two download methods. On the left, a laptop icon with a terminal window is labeled 'DTT'. On the right, a black box contains a JSON API response, labeled 'GDC API'. A red border encloses both the laptop and the API response.

```
{  
  "id": "c6bf94a6-9940-4155-86b4-bbb10875dbdb",  
  "data_format": "BEDPE",  
  "access": "controlled",  
  "file_name": "TCGA-BRCA_88cae21a-4890-4fdd-a678-c4864620942c.star_fusion.rna_fusion.bedpe",  
  "submitter_id": "f79ea5d2-6ad1-42c1-9928-1973c0e0413b",  
  "data_category": "Structural Variation",  
  "acl": [  
    "phs000178"  
  ],  
  "type": "structural_variation",  
  "file_size": 229,  
  "created_datetime": "2021-12-14T00:10:13.746858-06:00",  
  "md5sum": "6e5690795ff424264402ab9d2661b62b",  
  "updated_datetime": "2022-01-19T12:00:30.311996-06:00",  
  "file_id": "c6bf94a6-9940-4155-86b4-bbb10875dbdb",  
  "data_type": "Transcript Fusion",  
  "state": "released",  
  "experimental_strategy": "RNA-Seq",  
  "version": "1",  
  "data_release": "32.0 - 37.0"  
}
```

DTT

GDC API

This block provides a detailed view of the 'Access' section. It shows two radio button options: 'controlled' and 'open'. To the right, a table shows the number of files for each access type.

Access Type	# Files
<input type="checkbox"/> controlled	24
<input type="checkbox"/> open	4

Reminder: Download of controlled-access files requires dbGaP access and login through the GDC Data Portal.

Controlled Access Data

Controlled Data in the GDC

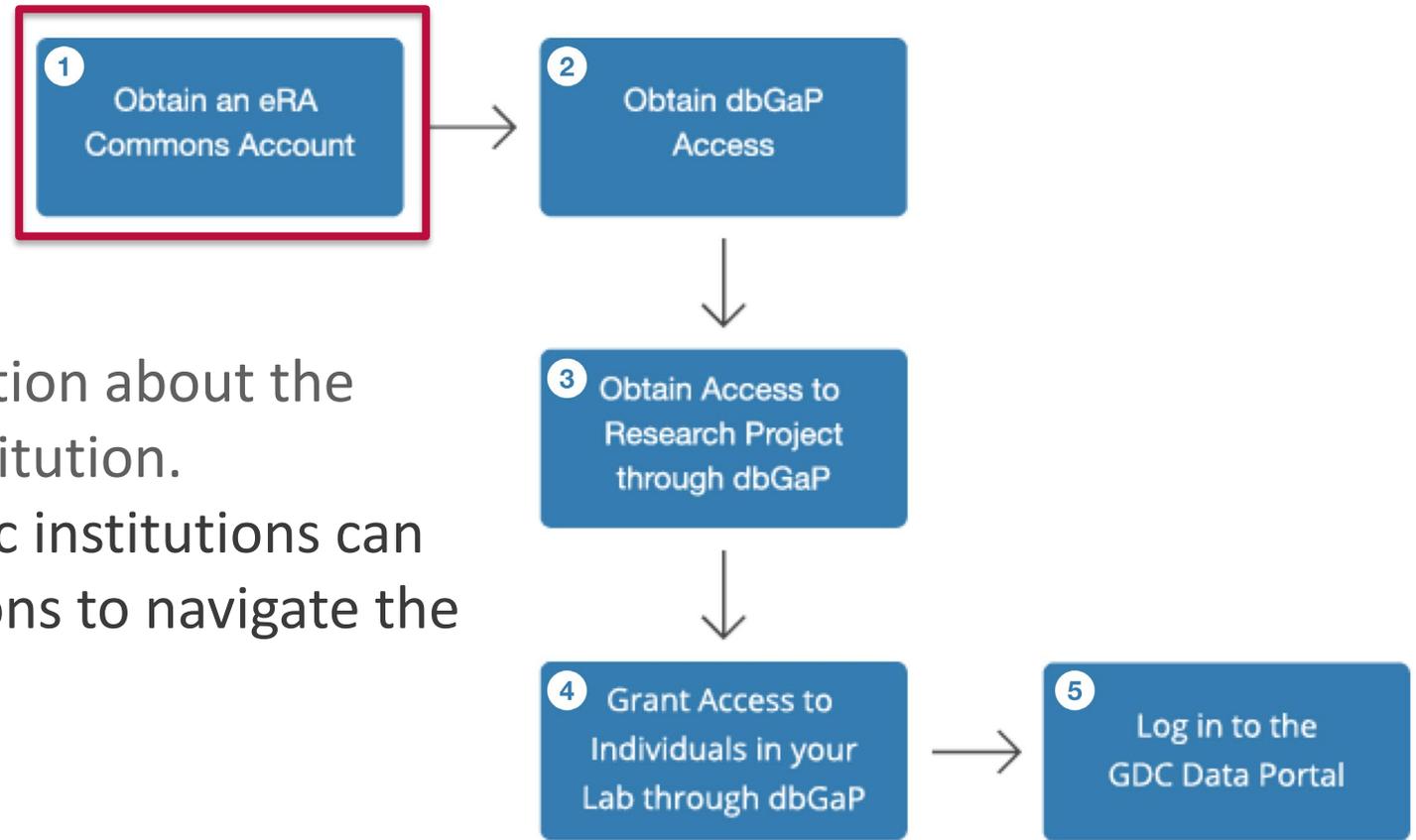
Files in the GDC are either open or controlled-access.

Controlled files usually contain identifiable information, such as genomic sequence

Open files may contain derived and much less identifiable information.

Exp. Strategy	Open	Controlled
Whole Exome	<ul style="list-style-type: none">Masked somatic mutation (MAF)	<ul style="list-style-type: none">Alignments (BAM)Unfiltered variants (MAF/VCF)
Whole Genome	<ul style="list-style-type: none">Copy number variation (TXT/TSV)	<ul style="list-style-type: none">Alignments (BAM)SSM variants (VCF)Structural variants (VCF/BEDPE)
RNA-Seq	<ul style="list-style-type: none">Gene expression (TSV)	<ul style="list-style-type: none">Alignments (BAM)Splice junction (TSV)Transcript fusion (BEDPE)
Methylation	All	N/A
Clinical / Biospecimen	All	N/A

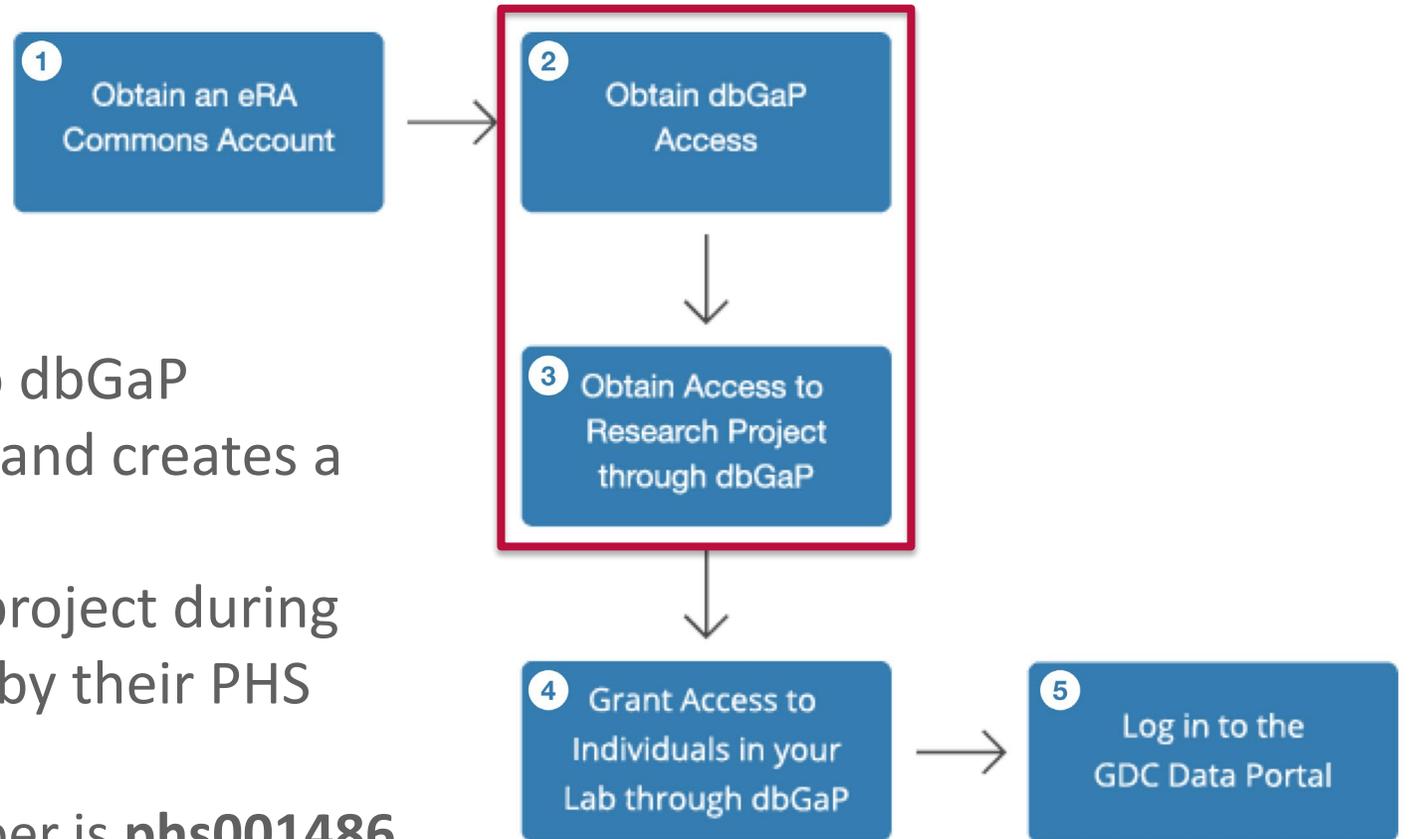
Obtaining Access to Controlled Data in the GDC



eRA Commons ID requires information about the users, their research, and their institution.

Researchers working with academic institutions can often get help from those institutions to navigate the eRA application process.

Obtaining Access to Controlled Data in the GDC

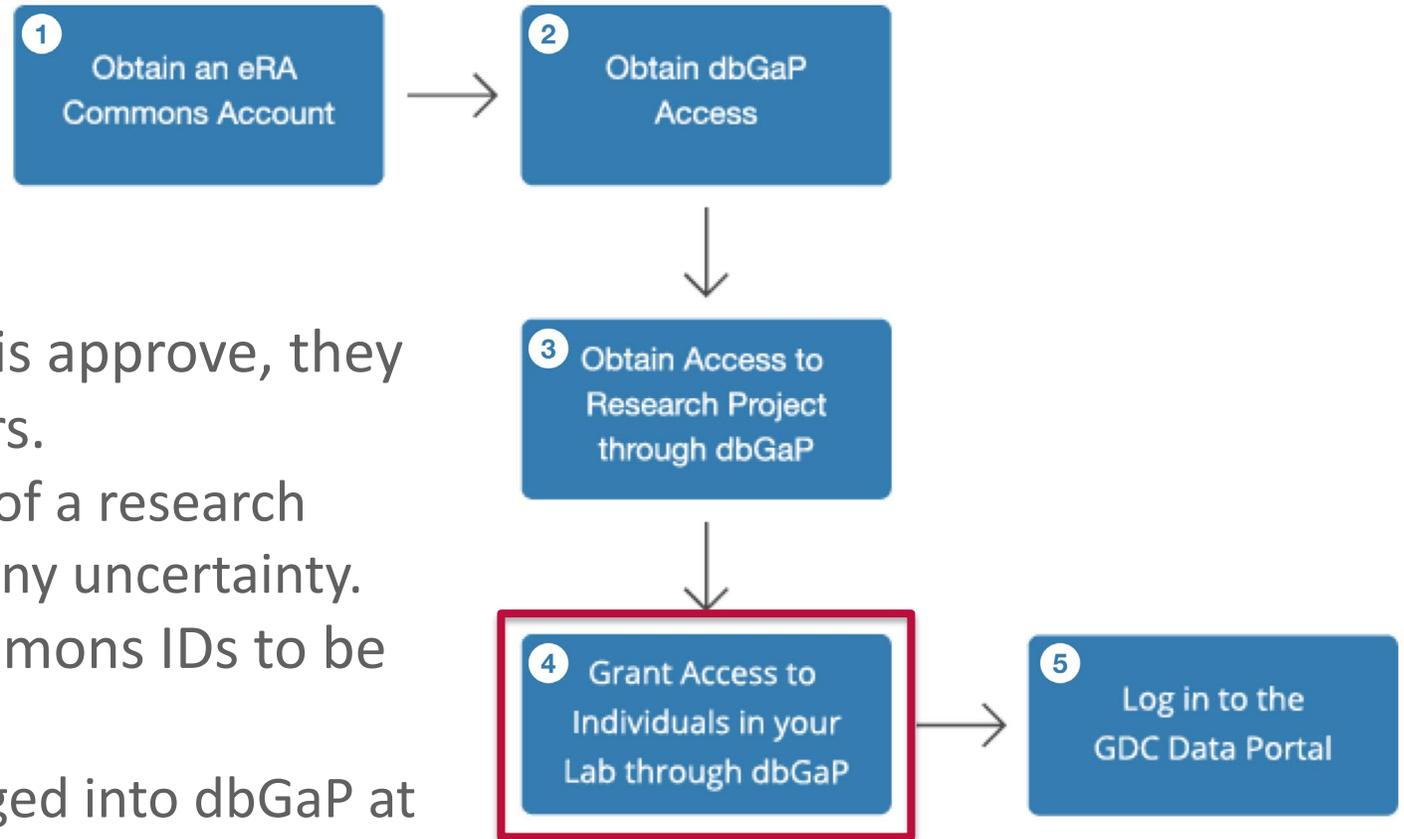


eRA Commons ID is used to log in to dbGaP
The PI of the group logs into dbGaP and creates a dbGaP project.

Specific data sets are added to the project during registration. Data sets are specified by their PHS number.

For the HCMI data set, this number is [phs001486](#)
Each data set has its own policies, which are outlined when the data set is added to the dbGaP project.

Obtaining Access to Controlled Data in the GDC

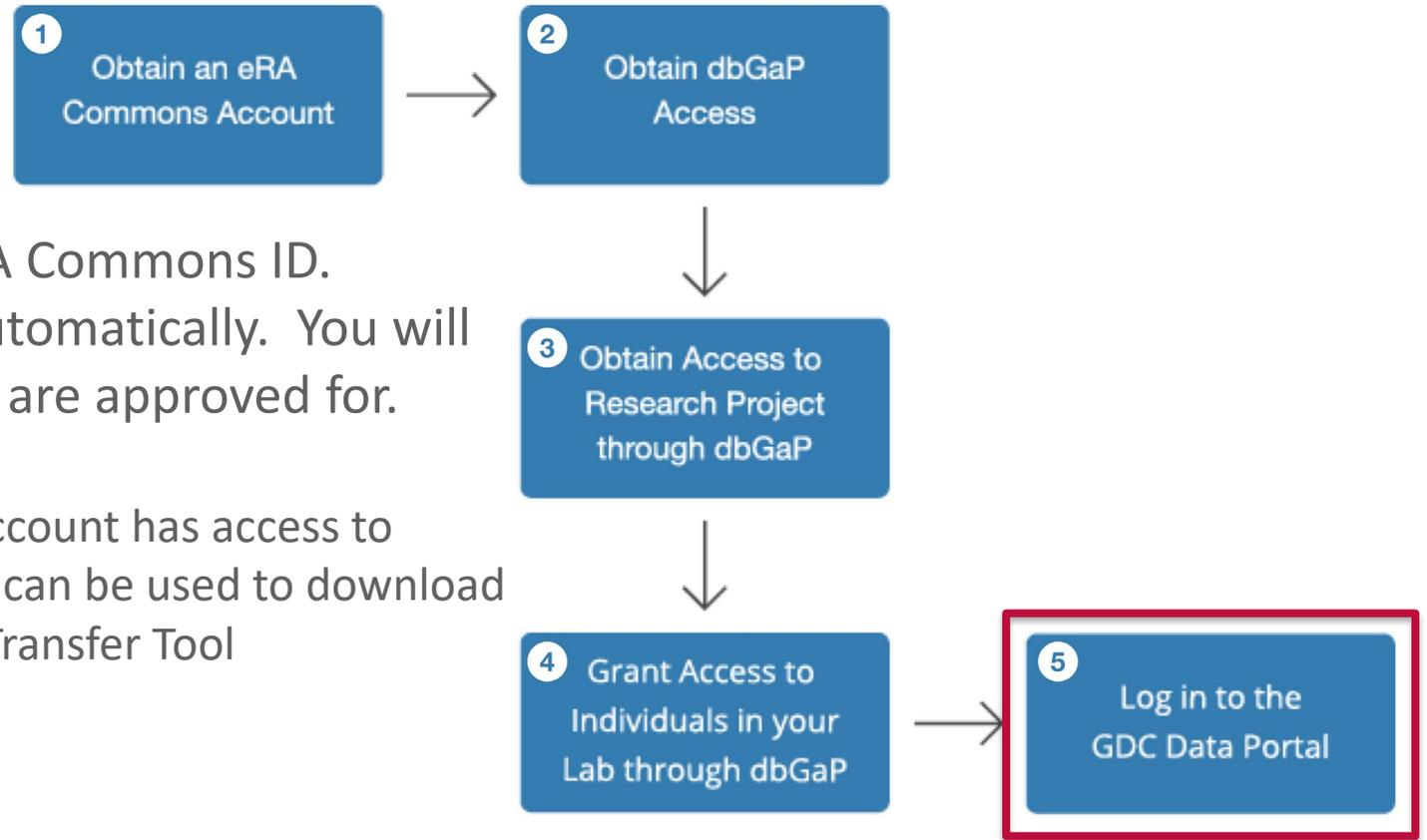


Once the PI's access to the data set is approved, they can provide access to team members.

Usually this applies to members of a research laboratory, contact dbGaP with any uncertainty. Team members must have eRA Commons IDs to be granted access.

Additionally, they must have logged into dbGaP at least once before.

Obtaining Access to Controlled Data in the GDC

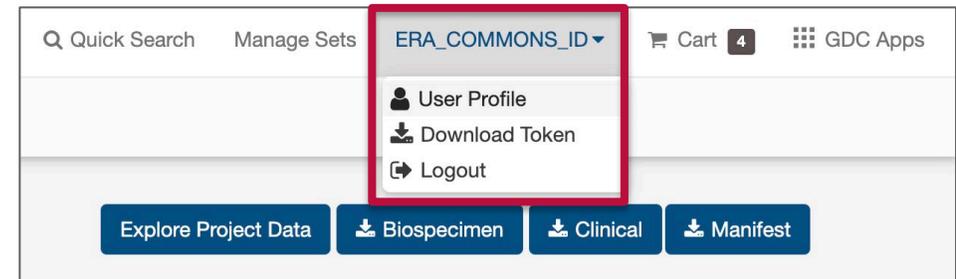
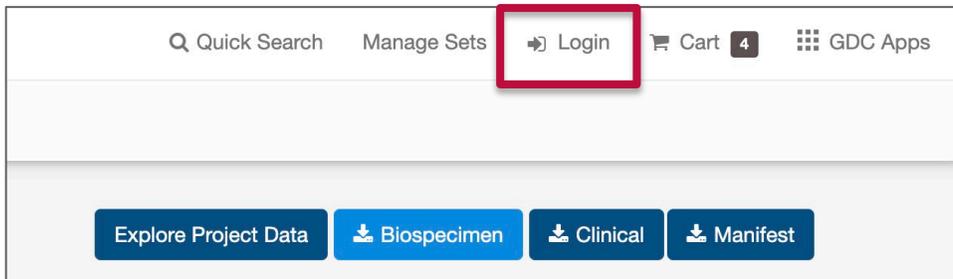


Log into the GDC Data Portal using the eRA Commons ID. Permissions from dbGaP will be applied automatically. You will have access to any controlled data set you are approved for.

Other Options:

User Profile: Shows which projects your account has access to

Download Token: Downloads a token that can be used to download controlled GDC data through API or Data Transfer Tool



Useful Links

GDC Portal – <https://portal.gdc.cancer.gov>

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis Repository Quick Search Manage Sets Login Cart 4 GDC Apps

Harmonized Cancer Datasets

Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

Q e.g. BRAF, Breast, TCGA-BLCA, TCGA-A5-A0G2

Data Portal Summary

[Data Release 37.0 - March 29, 2023](#)

PROJECTS 78	PRIMARY SITES 68	CASES 86,962
FILES 931,947	GENES 22,501	MUTATIONS 2,885,293

Cases by Major Primary Site

Adrenal Gland	1
Bile Duct	1
Bladder	2
Bone	1
Bone Marrow	9
Brain	2
Breast	9
Cervix	1
Colorectal	8
Esophagus	1
Eye	1
Head and Neck	3
Kidney	3
Liver	2
Lung	12
Lymph Nodes	1
Nervous System	4
Ovary	3
Pancreas	3
Pleura	1
Prostate	2
Skin	3
Soft Tissue	2
Stomach	2
Testis	1
Thymus	1
Thyroid	2
Uterus	3

Useful Links

GDC Documentation – <https://docs.gdc.cancer.gov>

Home API ▾ Data Portal ▾ Data Submission ▾ Data Transfer Tool ▾ Data Dictionary ▾ Data ▾ Encyclopedia 🔍

NCI Genomic Data Commons (GDC)
Documentation

*A place where researchers, data submitters and developers
can find detailed information on GDC processes and tools.*

🔍 What are you looking for?

Useful Links

GDC Website – <https://gdc.cancer.gov>

The screenshot shows the homepage of the National Cancer Institute Genomic Data Commons. At the top left is the NIH logo and the text "NATIONAL CANCER INSTITUTE Genomic Data Commons". To the right are links for "CCG Web Site", "Contact Us", "Launch Data Portal", and "GDC Apps". A search bar is located below these links. A navigation menu contains links for "About the GDC", "About the Data", "Analyze Data", "Access Data", "Submit Data", "For Developers", "Support", and "News". The main content area features a large section titled "The Next Generation Cancer Knowledge Base" with a sub-section "Cases by Major Primary Site" containing a colorful pie chart. To the right of the pie chart is text describing the GDC's role in providing a unified repository for cancer research. Further right, there are two highlighted sections: "Analyze Data" with a magnifying glass icon and "Access Data" with a DNA helix icon. Each section includes a brief description and a link for more information.

NIH NATIONAL CANCER INSTITUTE Genomic Data Commons

CCG Web Site | Contact Us | [Launch Data Portal](#) | GDC Apps

Search this website

/**/

About the GDC | About the Data | Analyze Data | Access Data | Submit Data | For Developers | Support | News

The Next Generation Cancer Knowledge Base

Cases by Major Primary Site

The NCI's Genomic Data Commons (GDC) provides the cancer research community with a unified repository and cancer knowledge base that enables data sharing across cancer genomic studies in support of precision medicine.

The GDC supports several cancer genome programs at the NCI Center for Cancer Genomics (CCG), including The Cancer Genome Atlas (TCGA) and

Analyze Data

The **GDC Data Analysis, Visualization, and Exploration (DAVE) Tools** allow users to interact intuitively with the GDC data and promote the development of a true cancer genomics knowledge base.

[More about Analyzing Data](#)

Access Data

The **GDC Data Portal** provides a platform for efficiently querying and

Useful Links

GDC Help Desk – support@nci-gdc.datacommons.io

The screenshot displays the National Cancer Institute Genomic Data Commons website. At the top left is the NIH logo and the text "NATIONAL CANCER INSTITUTE Genomic Data Commons". To the right are links for "CCG Web Site", "Contact Us", "Launch Data Portal", and "GDC Apps", along with a search bar. A navigation menu includes "About the GDC", "About the Data", "Analyze Data", "Access Data", "Submit Data", "For Developers", "Support", and "News". The "Support" link is highlighted with a red box. Below the navigation, a "Support" section is highlighted with a red box, containing links for "Monthly Support Webinar", "Help Desk: support@nci-gdc.datacommons.io", and "Join the GDC User Listserv". Other sections include "ANALYZE DATA", "ACCESS DATA", "SUBMIT DATA", and "FOR DEVELOPERS", each with a list of guides and FAQs. A "Twitter" link and "National Cancer Institute" text are also visible. The bottom of the page features a colorful pie chart, a section on "Access Data", and a description of the GDC's role in supporting cancer genome programs.

U.S. Department of Health & Human Services
National Institutes of Health | National Cancer Institute

<https://www.cancer.gov/>

1-800-4-CANCER

Produced July 2023

Help us prioritize future HCMI model releases

Hundreds of models from over a dozen different tissues have not yet entered our release pipeline:

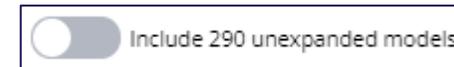
- Bladder
- Blood and lymphatic system
- Brain
- Breast
- Connective and soft tissue
- Esophagus
- Gynecological tissue
- Head and neck
- Kidney
- Liver
- Digestive tract
- Lung
- Pancreas
- Skin
- Stomach
- Thyroid

1. Browse and search these unreleased models at ATCC

- Use the “Submit your Input” button on the HCMI Landing page
 - Direct link: www.atcc.org/hcml-input

2. Explore them at the HCMI Searchable Catalog

- Look for the “unexpanded model” toggle near the top



3. Email us which models are most relevant for your research

- Contact us at: hcml@atcc.org

