

Download TCGA data from the GDC Data Portal as a GCT file

If you'd like a more comprehensive tutorial of the GDC website, you can find it here:
https://docs.gdc.cancer.gov/Data_Portal/Users_Guide/Getting_Started/

Download a Manifest and Metadata files from the GDC data portal. These two files contain sufficient information to download TCGA data from GDC.

1. Navigate to the GDC Data Portal

- Navigate to `gdc.cancer.gov`
- Click Launch Data Portal at the bottom of the homepage

The screenshot shows the GDC Data Portal homepage. At the top, there is a navigation bar with links: About the GDC, About the Data, Analyze Data, Access Data, Submit Data, For Developers, Support, and News. Below this is a large section titled "The Next Generation Cancer Knowledge Network" which includes a pie chart titled "Cases by Major Primary Site" and a "Data Availability Summary" table. A red arrow points from the "Launch Data Portal" button at the bottom of the page to the "Data Availability Summary" table. The "Launch Data Portal" button is highlighted with a red border.

Category	Count
Programs	3
Projects	40
Primary Sites	60
Cases	32,555

2. Type the name of the dataset

- Type the name of the dataset we are looking for "BRCA"
- Select the "TCGA-BRCA" dataset

Harmonized Cancer Datasets
Genomic Data Commons Data Portal

Get Started by Exploring:

Projects Exploration Analysis Repository

Q BRCA

PR TCGA-BRCA
TCGA-BRCA

GN BRCC3
BRCA1/BRCA2-containing complex, subunit 3

GN BRAT1
BRCA1-associated ATM activator 1

GN BRAP
BRCA1 associated protein

GN BCCIP
BRCA2 and CDKN1A interacting protein

310,859 22,147 3,142,246

Cases by Major Primary Site

Adrenal Gland
Bile Duct
Bladder
Blood
Bone
Bone Marrow
Brain
Breast
Cervix
Colon
Esophagus
Eye
Head and Neck
Kidney
Liver
Lung
Lymph Nodes
Nervous System
Ovary
Pancreas
Pituitary
Prostate
Skin
Soft Tissue
Stomach
Testis
Thyroid
Uterus

GDC Applications

The GDC Data Portal is a robust data-driven platform that allows cancer researchers and bioinformaticians to search and download cancer data for analysis. The GDC applications include:

Data Portal Website Data Transfer Tool API Data Submission Portal Documentation Legacy Archive

Site Home | Policies | Accessibility | FOIA
U.S. Department of Health and Human Services | National Institutes of Health | National Cancer Institute | USA.gov
NIH... Turning Discovery Into Health
UI @ 1.11.1, API 1.12.0 @ 4c0d76, Data Release 10 - December 21, 2017

3. Select data type

- In the "Experimental Strategy" panel, select RNA-Seq (click on the number of cases)

TCGA-BRCA

Download Biospecimen Download Clinical Download Manifest

Summary

Project ID TCGA-BRCA
Project Name Breast Invasive Carcinoma
Disease Type Breast Invasive Carcinoma
Primary Site Breast
Program TCGA

CASES 1,098
FILES 27,207
ANNOTATIONS 78

Cases and File Counts by Data Category

Data Category	Cases (n=1,098)	Files (n=27,207)
Raw Sequencing Data	1,098	4,604
Transcriptome Profiling	1,097	6,080
Simple Nucleotide Variation	1,044	8,648
Copy Number Variation	1,098	4,446
DNA Methylation	1,095	1,234
Clinical	1,097	1,097
Biospecimen	1,098	1,098

Cases and File Counts by Experimental Strategy

Experimental Strategy	Cases (n=1,098)	Files (n=27,207)
Genotyping Array	1,096	4,446
Methylation Array	1,095	1,234
WXS	1,050	10,823
RNA-Seq	1,092	4,898
miRNA-Seq	1,079	3,621

4. Navigate to files tab

- In the upper left corner, click on the files tab

The screenshot shows the National Cancer Institute GDC Data Portal interface. The 'Files' tab is selected in the top left navigation bar, highlighted with a red circle. A red arrow points from this tab to the 'Files' section of the main content area. The interface displays filters for Case, Case ID, Primary Site, Program, Project, Disease Type, and Gender. The main content area shows a table of files with columns for Case ID, Project, Primary Site, Gender, and various data categories like Seq, Exp, SNV, CNV, Meth, Clinical, Bio, and Annotations. The table shows 10 rows of data for TCGA-BRCA samples.

Case ID	Project	Primary Site	Gender	Files	Seq	Exp	SNV	CNV	Meth	Clinical	Bio	Annotations
TCGA-A8-A08S	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
TCGA-A8-A08F	TCGA-BRCA	Breast	Female	29	3	3	16	4	1	1	1	2
TCGA-A8-A084	TCGA-BRCA	Breast	Female	14	2	5	0	4	1	1	1	2
TCGA-A8-A09K	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
TCGA-A2-A0YT	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
TCGA-A8-A09F	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
TCGA-SL-AAT1	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
TCGA-A8-A07E	TCGA-BRCA	Breast	Female	14	2	5	0	4	1	1	1	1
TCGA-A7-A0CD	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
TCGA-BH-A0B6	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
TCGA-A0-A0J5	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
TCGA-AR-A2LR	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
TCGA-LL-A6FP	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1

5. Select HTSeq counts

- In the Workflow Type panel, click on "HTSeq - Counts"

The screenshot shows the GDC Data Portal interface. On the left, the 'Workflow Type' panel is expanded, and 'HTSeq - Counts' is selected, highlighted with a red circle and an arrow. The main panel displays a list of cases with columns for Case ID, Project, Primary Site, Gender, Files, and Available Files per Data Category. The 'HTSeq - Counts' workflow type is highlighted in the list.

6. Add files to Cart

6-a Add *all* files to cart

- Click the button "Add All Files to Cart"
- Then click on "Cart" in the top right corner

The screenshot shows the GDC Data Portal interface. The 'Add All Files to Cart' button is highlighted with a red circle and an arrow labeled '1'. The 'Cart' button in the top right corner is also highlighted with a red circle and an arrow labeled '2'. The main panel displays a list of cases with columns for Case ID, Project, Primary Site, Gender, Files, and Available Files per Data Category.

6-b Add some files to cart

- In the upper left corner of the center panel, click on the files tab

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis **Repository**

Quick Search Manage Sets Login Cart GDC Apps

Files Cases

Add a File Filter

File: e.g. 142682.bam, 4f6e2e7a-b...

Data Category: Transcriptome Profiling 1,222

Data Type: Gene Expression Quantification 1,222

Experimental Strategy: RNA-Seq 1,222

Workflow Type: HTSeq - Counts 1,222

Data Format: TXT 1,222

Platform: No data for this field

Access: open 1,222

Files (1,222) Cases (1,092)

Primary Site Project Disease Type Gender Vital Status

Showing 1 - 20 of 1,092 cases

Cart	Case UUID	Case ID	Project	Primary Site	Gender	Files	Seq	Exp	SNV	CNV	Meth	Clinical	Bio	Annotations
<input type="checkbox"/>	2779fa01-ac93-4e60-a997-338572172c3	TCGA-A8-A08S	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
<input type="checkbox"/>	6f959c92-b79f-4f84-9ee5-d07d3212d52d	TCGA-A8-A09K	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
<input type="checkbox"/>	3db99d1-e3b8-4ee2-b6a9-2a2e12c6fba9	TCGA-A8-A084	TCGA-BRCA	Breast	Female	14	2	5	0	4	1	1	1	2
<input type="checkbox"/>	0a017115-1c6b-45e7-8d55-e0a71df1b2e6	TCGA-A8-A08F	TCGA-BRCA	Breast	Female	29	3	3	16	4	1	1	1	2
<input type="checkbox"/>	dfebf76a-ec6b-4cd2-9d45-2e1e4befc7ea	TCGA-A8-A08E	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
<input type="checkbox"/>	14b95463-2108-4921-afc2-e29ee52b1bf	TCGA-A2-A0YT	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	2
<input type="checkbox"/>	09765b0a-94f6-47d2-af56-93368084ac3a	TCGA-A7-A0CD	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
<input type="checkbox"/>	853456d8-2367-404e-abc9-020fe8bbd209	TCGA-A8-A07E	TCGA-BRCA	Breast	Female	14	2	5	0	4	1	1	1	1
<input type="checkbox"/>	16fc3677-0393-4ed1-ad3f-c8355f056369	TCGA-SL-AAT1	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
<input type="checkbox"/>	03da7a65-fa2f-42cb-a709-5fe8736f1d3	TCGA-D8-A146	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
<input type="checkbox"/>	02f5aa33-a563-4ecb-9e33-dfa500a44931	TCGA-A8-A093	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1
<input type="checkbox"/>	1557183c-edcd-4df6-b98a-36546677fb69	TCGA-A8-A09J	TCGA-BRCA	Breast	Female	32	4	5	16	4	1	1	1	1

- Click on the shopping cart button (🛒) for all the files you want to download

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis **Repository**

Quick Search Manage Sets Login Cart GDC Apps

Files Cases

Add a File Filter

File: e.g. 142682.bam, 4f6e2e7a-b...

Data Category: Transcriptome Profiling 1,222

Data Type: Gene Expression Quantification 1,222

Experimental Strategy: RNA-Seq 1,222

Workflow Type: HTSeq - Counts 1,222

Data Format: TXT 1,222

Platform: No data for this field

Access: open 1,222

Files (1,222) Cases (1,092)

Primary Site Project Data Category Data Type Data Format

Showing 1 - 20 of 1,222 files

File Name	Cases	Project	Data Category	Data Format	File Size	Annotations
open 148d950b-4202-4f3f-ba15-84735cd08a48.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	256.34 KB	0
open a2233404-f380-4bd6-8770-0d4ba8988f69.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	250.03 KB	0
open e512f861-c59f-4272-b528-7424a882c795.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	254.02 KB	0
open 7388a1f1c58b-c4e79-83ef-d6871ef4cb6f.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	255.79 KB	0
open 9ad28472-32f0-4cb8-ba51-17d3e75fe85d.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	255.33 KB	0
open 3c49cb1a-c0f5-4b74-81f9-aa18d3da1e43.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	251.76 KB	0
open de2f8e0e-172-45e2-aa15-f543758deb78.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	255.79 KB	0
open 379553b2-13c2-4543-a9c3-9616fc249730.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	257.4 KB	0
open 0320c3aa-0b8c-4bde-854a-ee5a2d48eda.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	263.09 KB	0
open b7eefac-da30-4eed-9670-4afc45f86b4b.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	254.47 KB	0
open b3c6981a-b5db-481c-bc27-76a114856b03.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	250.92 KB	0
open 0c2cf133-50aa-44c8-b4a8-5d17ac2f5444.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	254.56 KB	0
open 3a611b3f-19bf-4f8a-ac15-770bcbad47c.htseq.counts.gz	1	TCGA-BRCA	Transcriptome Profiling	TXT	264.82 KB	0

- The files which have been added to your cart will change color

The screenshot shows the GDC Data Portal interface. The left sidebar has filters for File, Data Category, Data Type, Experimental Strategy, Workflow Type, Data Format, Platform, and Access. The main area displays a list of files with columns for Access, File Name, Cases, Project, Data Category, Data Format, File Size, and Annotations. A red arrow points to the 'Metadata' button in the 'Cart Items' section.

7. Download Metadata File

- Click on the "Metadata" button

The screenshot shows the GDC Data Portal interface. The left sidebar has filters for File, Data Category, Data Type, Experimental Strategy, Workflow Type, Data Format, Platform, and Access. The main area displays a list of files with columns for Access, File Name, Cases, Project, Data Category, Data Format, File Size, and Annotations. A red arrow points to the 'Metadata' button in the 'Cart Items' section.

8. Download Manifest File

- Hover your mouse over the "Download" button and click on "Manifest"

NIH NATIONAL CANCER INSTITUTE GDC Data Portal

Home Projects Exploration Analysis Repository

Quick Search Manage Sets Login Cart 1,222 GDC Apps

FILES 1222

CASES 1092

FILE SIZE 310.76 MB

File Counts by Project

Project	Cases (n=1,092)	Files (n=1,222)	File Size (Σ=310.76 MB)
TCGA-BRCA	1092	1,222	310.76 MB

File Counts by Authorization Level

Level	Files (n=1,222)	File Size (Σ=310.76 MB)
Authorized	1222	310.76 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.

Cart Items

Showing 1 - 20 of 1,222 files

Access	File Name	Cases Project	Data Category	Data Format	File Size	Annotations
open	0320d3aa-0b8c-4bda-854a-eed5a2d48eda.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	253.09 KB	0
open	b76eefac-da30-4ee4-9670-4afc45f86b4b.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	254.47 KB	0
open	b3b6981a-b5db-481c-bc27-76a114856b03.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	250.92 KB	0
open	148d950b-4202-49f1-be15-84735cd08a48.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	256.34 KB	0
open	a223340d-f980-4bde-8770-0d4ba898f0d9.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	250.03 KB	0
open	57b2e9d8-ddae-4457-9909-0d911288e8dc.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	253.01 KB	0
open	42d445c7-690c-4c0d-a268-ad3aef6f7776.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	254.16 KB	0
open	a3b9677e-a154-41f1-857c-4cd5244b77c3.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	259.55 KB	0
open	c0ad23b6-ec07-4756-a37e-b5439b4cc3d.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	256.34 KB	0
open	c1a570df-7a2f-44a1-9569-8a7eeb651d0.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	252.11 KB	0
open	f1afa28b-fac9-4e2e-b447-36097ef45a0b.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	259.03 KB	0
open	c758c950-68bc-4154-9d43-44339da29889.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	253.9 KB	0
open	ec57a9a3-90f1-4559-a241-d647a9cc37a8.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	248.81 KB	0
open	70029e42-3b29-46f5-908e-7cd203103277.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	257.95 KB	0
open	2bb5a0d3-aa66-4138-8bec-38c7c36c95fc.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	246.96 KB	0

Buttons: Sample Sheet, Metadata, Download (dropdown), Remove From Cart

Dropdown menu: Manifest, Cart