

# 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/](https://docs.gdc.cancer.gov/Data_Portal/Users_Guide/Getting_Started/)  
([https://docs.gdc.cancer.gov/Data\\_Portal/Users\\_Guide/Getting\\_Started/](https://docs.gdc.cancer.gov/Data_Portal/Users_Guide/Getting_Started/))

## Download a Manifest and Metadata files from the GDC data portal

FAQ: What is a manifest file?

### 1. Navigate to the GDC Data Portal

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

Secure | <https://gdc.cancer.gov>

NIH NATIONAL CANCER INSTITUTE Genomic Data Commons

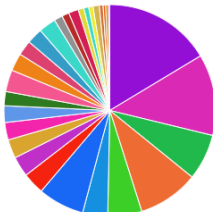
CCG Web Site | Contact Us | [Launch Data Portal](#) | GDC Apps

Search this website

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

### The Next Generation Cancer Knowledge Network

Cases by Major Primary Site



The NCI's Genomic Data Commons (GDC) provides the cancer research community with a unified data repository 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 Therapeutically Applicable Research to Generate Effective Treatments (TARGET).

→ More about the GDC

### High Quality Data Sharing Enables Precision Medicine

The GDC obtains validated datasets from NCI programs in which the strategies for tissue collection couples quantity with high quality.

The GDC encourages data sharing in support of precision medicine. Tools are provided to guide data submissions by researchers and institutions.

### 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 downloading high quality and complete data. The GDC also provides a **GDC Data Transfer Tool** and a **GDC API** for programmatic access.

→ More about Accessing Data

### Submit Data

The GDC provides tools to guide data submission including the **GDC Data Submission Portal**, a web-based tool for submitting clinical, biospecimen and small volumes of molecular data as well as the **GDC Data Transfer Tool**, a client-based tool for submitting large, high

### Data Availability Summary

Programs	3
Projects	40
Primary Sites	60
Cases	32,555

[Launch Data Portal](#)

## 2. Type the name of the dataset

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

The screenshot shows the GDC Data Portal search results for 'BRCA'. The search bar at the top left contains 'BRCA'. Below the search bar, the results are listed under the 'TCGA-BRCA' project. The results include the following data categories and counts:

Data Category	Cases	Files
BRCA1/BRCA2-containing complex, subunit 3	310,859	22,147
BRCA1-associated ATM activator 1	310,859	22,147
BRCA1-associated protein	310,859	22,147
BRCA2 and CDKN1A interacting protein	310,859	22,147

On the right side of the page, there is a bar chart titled 'Cases by Major Primary Site' showing the distribution of cases across various cancer types. The chart shows that Breast is the most common primary site, followed by Lung, Prostate, and others.

## 3. Select data type

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

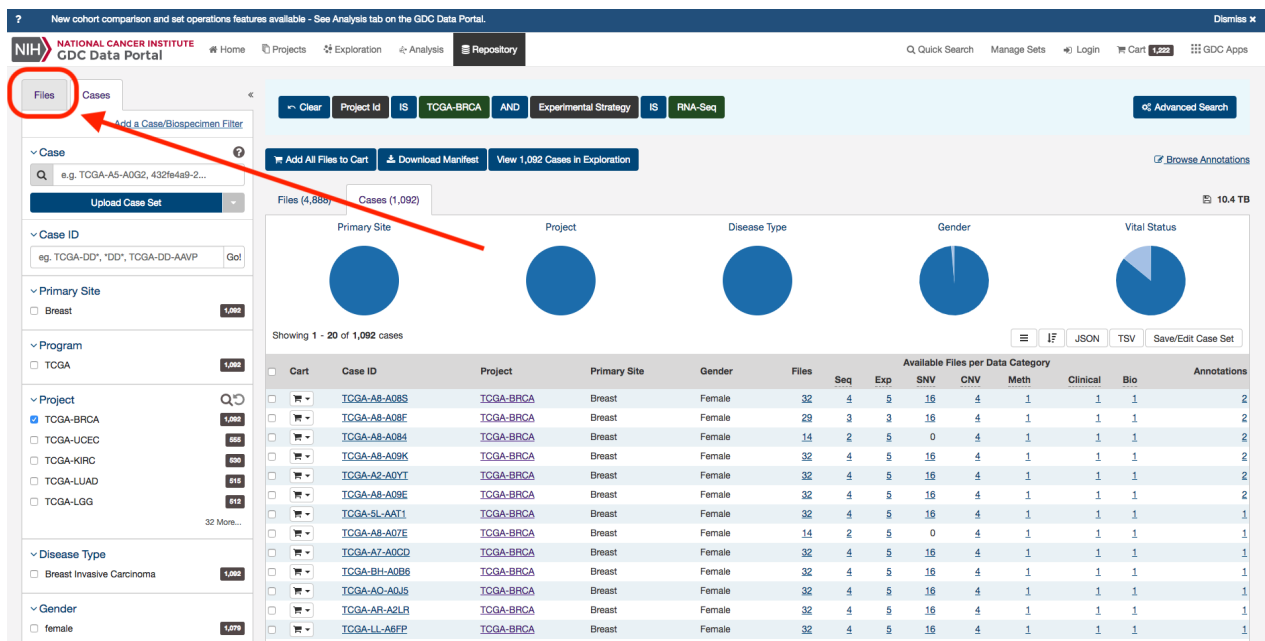
The screenshot shows the 'TCGA-BRCA' dataset page. The page displays a summary of the dataset, including the Project ID (TCGA-BRCA), Project Name (Breast Invasive Carcinoma), Disease Type (Breast Invasive Carcinoma), Primary Site (Breast), and Program (TCGA). The page also shows the number of cases (1,098) and files (27,207).

Below the summary, there are two tables showing the distribution of data by category and experimental strategy. The 'Cases and File Counts by Experimental Strategy' table is highlighted with a red box, and the 'RNA-Seq' row is selected.

Experimental Strategy	Cases	Files
Genotyping Array	1,098	4,446
Methylation Array	1,098	1,234
WXS	1,050	10,823
RNA-Seq	1,092	4,888
miRNA-Seq	1,079	3,621

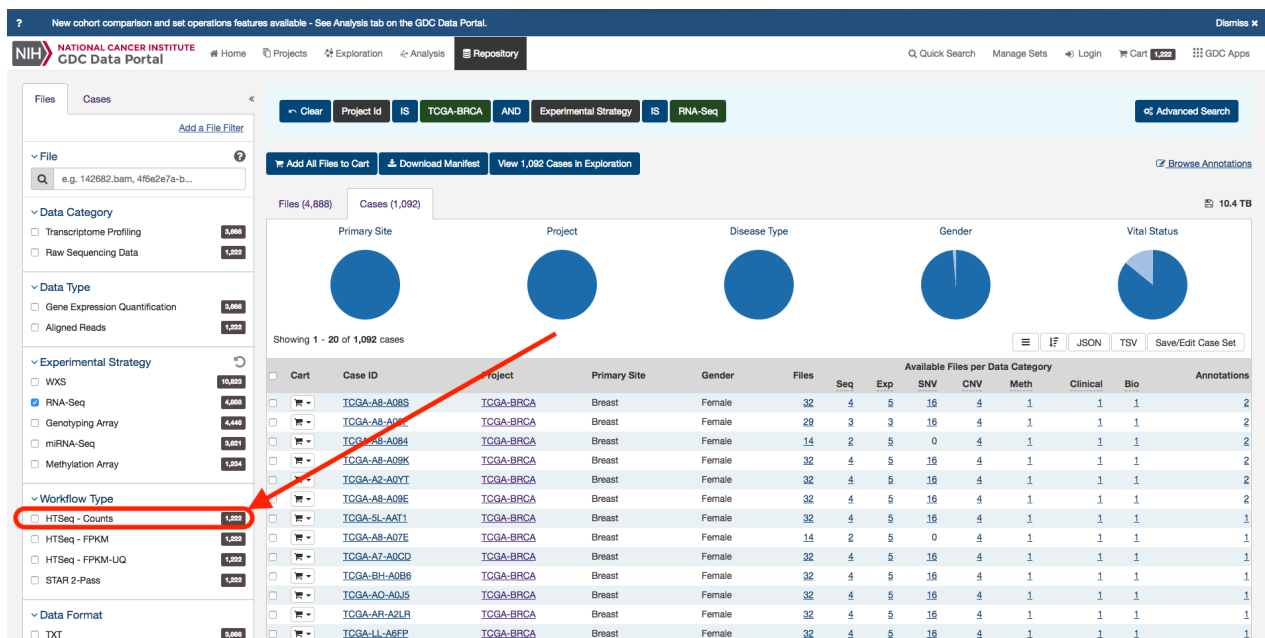
## 4. Navigate to files tab

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



## 5. Select HTSeq counts

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



## 6. Add files to Cart

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

The screenshot shows the TCGA Data Portal interface. At the top, there's a navigation bar with 'Home', 'Projects', 'Exploration', 'Analysis', and 'Repository'. The 'Repository' tab is active. On the right, there's a 'Cart' button with '1,222' items, highlighted by a red circle and an arrow. Below the navigation bar, there's a 'Files' section with a search bar and filters for 'Data Category', 'Data Type', 'Experimental Strategy', 'Workflow Type', 'Data Format', 'Platform', and 'Access'. The 'Data Category' filter is set to 'Transcriptome Profiling'. The 'Data Type' filter is set to 'Gene Expression Quantification'. The 'Experimental Strategy' filter is set to 'RNA-Seq'. The 'Workflow Type' filter is set to 'HTSeq - Counts'. The 'Data Format' filter is set to 'TXT'. The 'Platform' filter is set to 'No data for this field'. The 'Access' filter is set to 'open'. The 'Files' section shows '1,222' files and '1,092' cases. Below this, there are five pie charts representing 'Primary Site', 'Project', 'Disease Type', 'Gender', and 'Vital Status'. The 'Primary Site' chart is highlighted by a red circle and an arrow. Below the charts, there's a table showing 'Showing 1 - 20 of 1,092 cases'. The table has columns for 'Case ID', 'Project', 'Primary Site', 'Gender', 'Files', 'Seq', 'Exp', 'SNV', 'CNV', 'Meth', 'Clinical', 'Bio', and 'Annotations'. The first row of the table is 'TCGA-A8-A08S', 'TCGA-BRCA', 'Breast', 'Female', '32', '4', '5', '16', '4', '1', '1', '1', '2'.

## 7. Download Metadata File

- Click on the "Metadata" button

The screenshot shows the TCGA Data Portal interface. At the top, there's a navigation bar with 'Home', 'Projects', 'Exploration', 'Analysis', and 'Repository'. The 'Repository' tab is active. On the right, there's a 'Cart' button with '1,222' items. Below the navigation bar, there's a 'Files' section with a search bar and filters for 'Data Category', 'Data Type', 'Experimental Strategy', 'Workflow Type', 'Data Format', 'Platform', and 'Access'. The 'Data Category' filter is set to 'Transcriptome Profiling'. The 'Data Type' filter is set to 'Gene Expression Quantification'. The 'Experimental Strategy' filter is set to 'RNA-Seq'. The 'Workflow Type' filter is set to 'HTSeq - Counts'. The 'Data Format' filter is set to 'TXT'. The 'Platform' filter is set to 'No data for this field'. The 'Access' filter is set to 'open'. The 'Files' section shows '1,222' files and '1,092' cases. Below this, there are five pie charts representing 'Primary Site', 'Project', 'Disease Type', 'Gender', and 'Vital Status'. The 'Primary Site' chart is highlighted by a red circle and an arrow. Below the charts, there's a table showing 'Showing 1 - 20 of 1,092 cases'. The table has columns for 'Case ID', 'Project', 'Primary Site', 'Gender', 'Files', 'Seq', 'Exp', 'SNV', 'CNV', 'Meth', 'Clinical', 'Bio', and 'Annotations'. The first row of the table is 'TCGA-A8-A08S', 'TCGA-BRCA', 'Breast', 'Female', '32', '4', '5', '16', '4', '1', '1', '1', '2'. Below the table, there's a 'Cart Items' section showing 'Showing 1 - 20 of 1,222 files'. The table has columns for 'Access', 'File Name', 'Cases Project', 'Data Category', 'Data Format', 'File Size', and 'Annotations'. The first row of the table is 'open', '0320d3aa-0b6c-4bda-854a-eed5a2d48eda.htseq.counts.gz', '1 TCGA-BRCA', 'Transcriptome Profiling', 'TXT', '253.09 KB', '0'. Below the table, there's a 'Download Manifest' button, a 'Download Cart' button, and a 'Sample Sh...' button. The 'Metadata' button is highlighted by a red circle and an arrow.

## 8. Download Manifest File

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

?

New cohort comparison and set operations features available - See Analysis tab on the GDC Data Portal.

Dismiss

NIH

NATIONAL CANCER INSTITUTE

GDC Data Portal

Home

Projects

Exploration

Analysis

Repository

Quick Search

Manage Sets

Login

Cart 0,222

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-0b9c-4bda-854a-ee05a2d49eda.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	253.09 KB	0
open	b78eefac-da30-4eed-9670-4afc45f86b4b.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	254.47 KB	0
open	b3b6981a-b6db-481c-bc27-76a114856b03.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	250.92 KB	0
open	148d950b-4202-4f3f-be15-8473cd08a468.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	256.34 KB	0
open	a2233404-f380-4bd9-8770-0d4ba8998fd9.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	250.03 KB	0
open	57b2e808-d3ae-4457-9959-0d911288e8dc.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	253.01 KB	0
open	42d4f47-690c-4cd1-a2e8-ad3aef977778.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	256.58 KB	0
open	c0ae23b6-ec07-4756-a37e-bf5439b4cc3d.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	256.34 KB	0
open	c1a570cf-7a5f-44a1-9569-8a7eeeb651d0.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	252.11 KB	0
open	f1afa26b-fac9-4e2e-b447-36097ef45a0b.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	259.03 KB	0
open	c75bcf60-68bc-4154-9d43-44339da29889.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	253.9 KB	0
open	ec57a9a3-90f1-455b-a241-d647a9ca37a6.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	248.81 KB	0
open	70029e42-3bc9-48f5-908e-7cd203103277.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	257.95 KB	0
open	2bb5a0f3-eeae-4138-8bec-38c7c35c95fc.htseq.counts.gz	1 TCGA-BRCA	Transcriptome Profiling	TXT	246.96 KB	0