



# Load Data

## Overview

Navigating GDC's Data Portal

### Step 1: Move the Manifest File to a Clear Location

- 1 Move the Manifest File to a Clear Location
- 2 Move the GDC Client to a Safe Location

### Step 2: Downloading the Data

- 1 Navigate to Your TCGA Folder
- 2 Start the Download

### Troubleshooting

- 1 Being vigilant of file size
- 2 If you are blocked from opening/downloading the files

## Overview

1. Set up GDC Client correctly on my Mac.
2. Move the manifest file to an easy-to-use location.
3. Download the data using the GDC Client.

## Navigating GDC's Data Portal

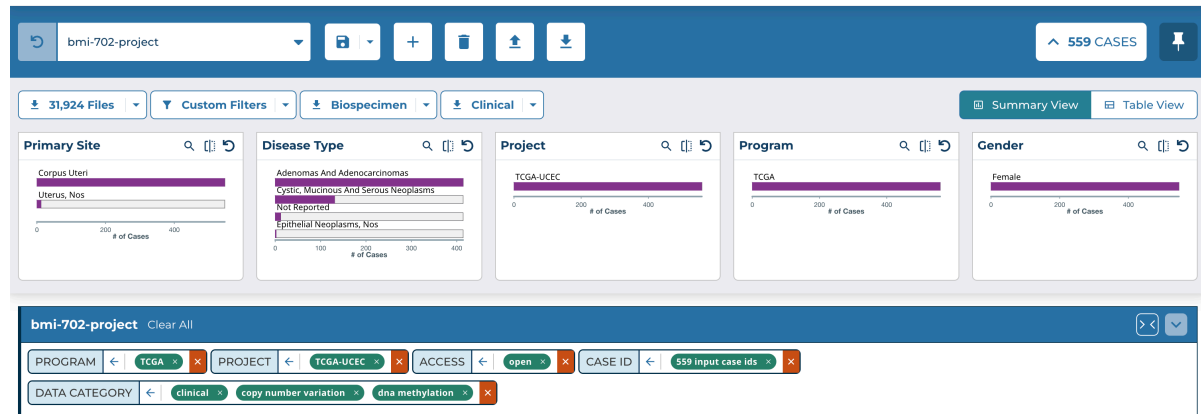
1. Link to data: <https://portal.gdc.cancer.gov/>
2. Our project's analysis focuses requires these filters:

### Program/Project

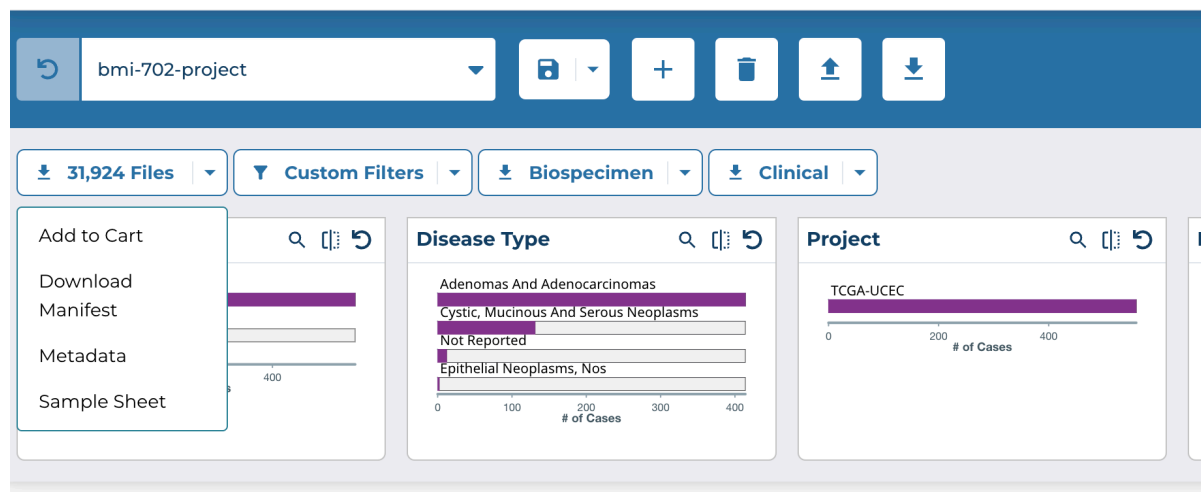
- ☐ TCGA - UCEC

### Data Categories:

- ☐ clinical
- ☐ copy number variation
- ☐ dna methylation










### 3. Save manifest file



### 4. Go to GDC Data Transfer Tool (<https://gdc.cancer.gov/access-data/gdc-data-transfer-tool>)

### 5. Download the Tool client (depending on the computer you have)

## Binary Distributions

Platform	GDC Data Transfer Tool Client	GDC Data Transfer Tool UI
	<a href="#">Release Notes</a>   <a href="#">User's Guide</a>	<a href="#">Release Notes</a>   <a href="#">User's Guide</a>
MAC OSX x64 (Intel Processor)	 <a href="#">gdc-client_2.3.0_OSX_x64-py3.8-macos-12</a>	 <a href="#">gdc_dtt-ui_v1.0.0_mac_x64.zip</a>
MAC OSX x64 (Silicon Processor)	 <a href="#">gdc-client_2.3.0_OSX_x64-py3.8-macos-14</a>	--
Windows x64	 <a href="#">gdc-client_2.3.0_Windows_x64-py3.8-windows-2019</a>	 <a href="#">gdc_dtt-ui_v1.0.0_win_x64.zip</a>
Ubuntu x64	 <a href="#">gdc-client_2.3.0_Ubuntu_x64-py3.8-ubuntu-20.04</a>	 <a href="#">gdc_dtt-ui_v1.0.0_linux_x86_64.AppImage.zip</a>

## Step 1: Move the Manifest File to a Clear Location

### 1 Move the Manifest File to a Clear Location

open terminal and run:

```
mkdir -p ~/Documents/TCGA
mv ~/Downloads/bmi-702-ucec-manifest.txt ~/Documents/TCGA/manifest.txt
```

#### Code Explanation:

I named the manifest file as `bmi-702-ucec-manifest.txt` and put it in my Downloads.

I created a folder called `TCGA` in the `Documents` folder on my computer

I renamed `bmi-702-ucec-manifest.txt` to `manifest.txt` and moved it to `~/Documents/TCGA`

### 2 Move the GDC Client to a Safe Location

Run this:

```
sudo mv ~/Downloads/gdc-client /usr/local/bin/
```

after running, You will need to enter your Mac password.

### Code Explanation:

My device has a MAC OSX 64 Silicon processor. Once I downloaded the relevant zip file from the GDC Data Transfer Tool, it saved to my downloads @ `/Users/nehadantuluri/Downloads/gdc-client`

Then, with this code, I move it to a system folder where we can run it easily from anywhere.

This moves `gdc-client` to `/usr/local/bin/`, so you can now run it from any directory.

---

## Step 2: Downloading the Data

### 1 Navigate to Your TCGA Folder

To ensure you are in the right folder before downloading, run this in terminal

```
cd ~/Documents/TCGA/
```

### 2 Start the Download

To download the files, type:

```
gdc-client download -m manifest.txt --n-processes 2
```

### Code Explanation:

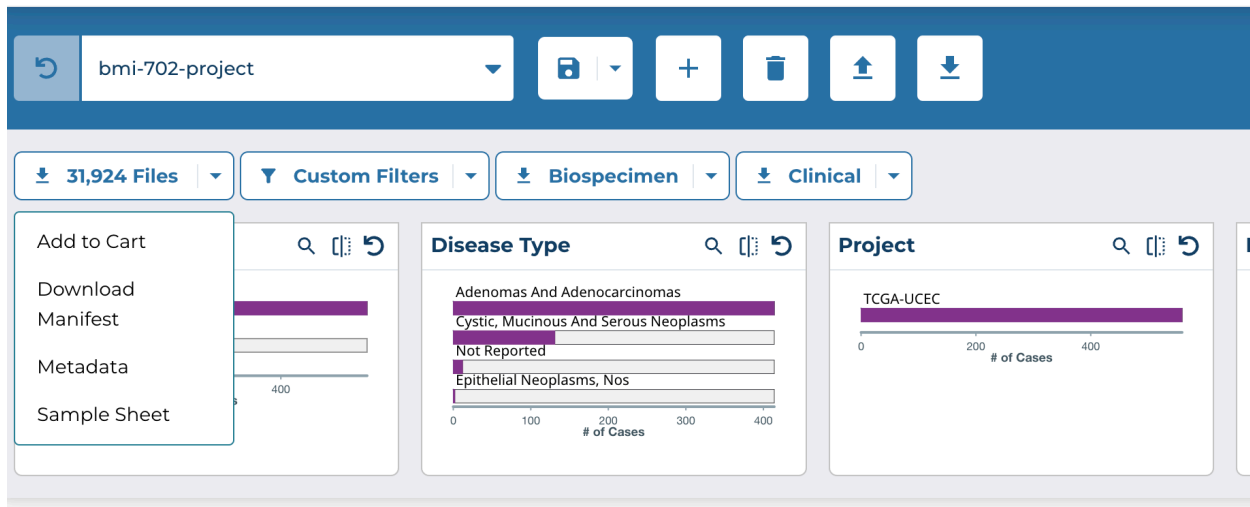
- Download only the files listed in your manifest file.
- Use 2 download streams at a time to avoid overloading your computer.

---

## Troubleshooting

### 1 Being vigilant of file size

In the example below, there are 31,924 files. Your computer will not have the space to load this. Use the cohort builder to filter and take the data you need only. For example, we took clinical data only, then gene methylation data only, etc.



## 2 If you are blocked from opening/downloading the files

Open "System Settings" and press Privacy & Security. Then, scroll down to **Security and wither** look for a popup message or press on the lock to make changes to your existing settings. Once you are here, be sure to press "allow anyway"

Here are some pop-up images you may get:

