



General Commons

GENERAL COMMONS (GC)

User Guide

Contents

Introduction.....	2
Overview.....	2
Programs	3
Studies	3
GC Home Page	4
GC Data Page	5
GC Cart Page.....	9
How to use GraphQL interface:	11

Introduction

The purpose of this document is to help users navigate the General Commons (GC) data portal and provide instructions on accessing and analyzing research data shared by the GC.

Overview

The GC is one of several data commons within the [Cancer Research Data Commons \(CRDC\)](#) infrastructure for storing and sharing cancer research data generated by NCI-funded programs. GC currently hosts a variety of data types from NCI projects and programs such as the Human Tumor Atlas Network (HTAN), Division of Cancer Control and Population Sciences (DCCPS), and Childhood Cancer Data Initiative (CCDI), as well as data from independent research projects.

The GC provides data-sharing capabilities for additional studies that fall under the following categories:

- Studies with data that do not match any existing CRDC data commons
- Studies with data that do not fit current data type criteria for any CRDC data commons

Searching the data housed in the GC is easy to do through the portal's Data Explore page, which has a filter function to find data by categories. Users can build cohorts from metadata on the Data Explore page and export those cohorts to Seven Bridges- CGC (SB-CGC), from Velsera to access the files for further analysis in this NCI-funded secure cloud environment, without the expense of downloading data to a local compute environment.

The metadata on the GC is searchable by all and can build cohorts to export to Velsera. Access to the controlled data files. Users requires approval from the NIH Data Access Committee through the [dbGaP](#) process.

Open-access data is publicly accessible and does not require any approvals. Being a cloud repository, the GC does not facilitate direct downloads of data, owing to high data download charges.

Access to data in the GC for review and/or analysis is through the NCI's Cloud Resource, Velsera Cancer Genomics Cloud (Velsera/CGC) (formerly Seven Bridges). See below for a description of how to select data and import to Velsera/CGC. More instructions on how to access and upload files can be found in the section below or on the cart page.

Explore Data on the GC

Start on the GC Portal home page to learn how to navigate through the various pages and searches.

Programs

The Programs page on the General Commons (GC) platform is a dedicated section that organizes and presents cancer-related research programs or initiatives. This page allows users to explore various cancer research efforts, and the data associated with them. It's designed to provide an overview of the programs and the studies or projects they support, offering insights into different cancer research areas.

Studies

The Studies page on the GC platform typically serves as a centralized place where users can access information about various cancer research studies. It provides details on studies that have been submitted under different programs or initiatives within the GC. The page is designed to help users easily find studies related to specific dbGaP accessions.

GC Home Page

Below is the GC home page. A description of its features follows the image.

The GC home page.

NIH NATIONAL CANCER INSTITUTE
General Commons

SEARCH CDS

HOME DATA PROGRAMS STUDIES ABOUT 0 Files

EXPLORE GENERAL COMMONS

STUDIES 31 PARTICIPANTS 60641 SAMPLES 97022 FILES 375241

ABOUT THE GENERAL COMMONS

The General Commons (GC) is a data repository under the [Cancer Research Data Commons \(CRDC\)](#) infrastructure for storing cancer research data generated by NCI funded programs. The GC provides secure and authorized storage and data sharing capabilities on the cloud via the NCI's Cloud Resources for studies that are not a match for other CRDC Data Commons.

EXPLORE

STUDIES

View summaries of the studies submitted to GC.

GO TO STUDIES LISTING

SUBMIT DATA

To get started with a submission please submit a request on the CRDC Submission Portal.

CRDC HELPDESK

GC DATA

Search all GC data and build cohorts of participants and samples from the studies hosted in the GC. The data files pertaining to these cohorts can then be accessed and analyzed on a NCI Cloud Resource platform.

LEARN MORE

1. **HOME** – this page to generally explore the GC.
2. **DATA** – This tab opens the Data page where users can search studies using various filters.
3. **PROGRAMS** – This tab opens the Program listing page, where users can learn more about the programs that submit data to the GC, as well as associated studies within the program.
4. **STUDIES** – This tab opens the studies listing page where the user can see the studies that have submitted data to the GC
5. **ABOUT** – The About tab is a drop-down list that opens several pages with information about the GC, including:
 - A link to the submission requests
 - GraphQL API interface
 - Data Model
 - GC data and software releases

1. Stats bar across the top of the page

- **STUDIES** – A count of all the studies in the GC
- **PARTICIPANTS** – A count of all the participants in the GC
- **SAMPLES** – A count of all the samples in GC
- **FILES** – A count of all the files in GC

GC Data Page

The Data page contains all the metadata submitted along with each study. This is users can use the GC Data page to search for metadata and build cohorts, which they can select and add to the cart. The selected files can then be exported to the CGC, where users can access the action data files for further exploration and analysis. A description of the features and image of the Data page are below.

The GC Data Page

NIH NATIONAL CANCER INSTITUTE General Commons

SEARCH CDS

HOME DATA PROGRAMS STUDIES ABOUT

31 STUDIES 60641 PARTICIPANTS 97022 SAMPLES 375241 FILES

Demographics Study Samples Genomic Imaging Diagnosis Files

Experimental Strategy: 166671 Files (WXS)

Sex: 31 Participants (Decline to answer)

File Type: 5 Files (ADF)

Study Data Types: 67566 Files (Imaging/Genomic)

Image Modality: 352651 Files (Not specified in data)

Sample Type: 35 Samples (Abdomen, NOS)

Participants (60641) Samples (97022) Files (375241)

ADD ALL FILES ADD SELECTED FILES

ROWS PER PAGE: 10 1-10 OF 60641

Participant ID	Study Name	Accession	Sex	Samples
00301d78915737fa100f	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Female	7219_C71.9 : Brain, NOS, 7220_C42.0 : Blood
0061cbb0846973206fcf	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Male	28764_C06.0 : Buccal mucosa, 28896_C42.1 : Bone marrow
0065af91e89ee2859595	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Female	14815_C71.9 : Brain, NOS, 14897_C42.0 : Blood
008b04a84717e7d007a4	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Male	26023_C06.0 : Buccal mucosa, 26105_C76.2 : Abdomen, NOS, 29168_C22.0 : Liver, 30208_C17.9 : Small intestine, NOS
009f9efaedc9602a28fb	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Female	16704_C76.2 : Abdomen, NOS, 16705_C06.0 : Buccal mucosa, 18236_C76.2 : Abdomen, NOS, 18237_C80.9 : Unknown primary site
00c5e4372375eb3627f3	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Female	11772_C06.0 : Buccal mucosa, 11835_C71.9 : Brain, NOS, 12257_C71.9 : Brain, NOS
00df938d509984eada22	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Male	5903_C76.0 : Head, face or neck, NOS, 5922_C42.0 : Blood
00e3d1d383c8c08a25d2	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Male	3737_C76.2 : Abdomen, NOS, 3773_C42.0 : Blood
00eff50ac98a200da9f9	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	Male	28644_C42.1 : Bone marrow, 28664_C06.0 : Buccal mucosa

The following data filters are listed on the left side of the page.

Note: GC is file-centric, so the counts shown for each filter also represent the number of files.

- **STUDY** – If this study included controlled access data, the title will correspond to the dbGaP Study Name
- **EXPERIMENTAL STRATEGY** – The type of study or experimental data generated for testing or research purposes.
- **SAMPLE TUMOR STATUS** – Normal or Tumor Sample Pathology Indicator
- **SAMPLE TYPE** - Text term to describe the source of a biospecimen used for a laboratory test
- **SEX** – A textual description of a person's sex at birth

- **FILE TYPE** – A defined organization or layout representing and structuring data in a computer file.
- **PHS ACCESSION** – dbGaP study accession number
- **STUDY DATA TYPES** – The type of study or experimental data generated for testing or research purposes
- **IMAGE MODALITY** - The method in which the images are generated
- **LIBRARY STRATEGY** – The overall strategy for the collection of double-stranded DNA fragments flanked by oligonucleotide sequence adapters to enable their analysis by high-throughput sequencing.
- **LIBRARY SOURCE** – The source of a sample collection of double-stranded DNA fragments analyzed by high-throughput sequencing.
- **LIBRARY SELECTION** – The type of systematic actions performed to select or enrich DNA fragments used in analysis by high-throughput sequencing.
- **LIBRARY LAYOUT** – The read strategy or method that was used for sequencing and analysis of a nucleotide library.
- **PLATFORM** – The words used to describe the instrument used to carry out a high-throughput sequencing experiment.
- **INSTRUMENT MODEL** – The words used to describe the specific model of the instrument used to carry out a high-throughput sequencing experiment.
- **REFERENCE GENOME ASSEMBLY** – One or more characters are used to identify the published NCBI genetic sequence that is used as a reference against which other sequences are compared.
- **PRIMARY DIAGNOSIS** – Text term used to describe the patient's histologic diagnosis, as described by the World Health Organization's (WHO) International Classification of Diseases for Oncology (ICD-O)
- **NUM OF STUDY PARTICIPANTS** – Number of participants in the selected study/studies
- **NUM OF STUDY SAMPLES** – Number of samples in the selected study/studies

2. Donuts (the four circles – on the top half of the screen)

- **Experimental Strategy** – a visual representation of the various types of experimental strategies in the GC Portal
- **Sex** – visual representation of the files for sex in the portal
- **File Type** – a visual representation of all the count of file types in GC

- **Study Data Types** – a visual representation of the count of all the study data types in GC
- **Image Modality** – The method in which the images are generated
- **Sample Type** – Text term to describe the source of a biospecimen used for a laboratory test

3. Data Table (the light blue header across the top of the screen)

- **Participants** – The participants tab contains the number of participants in the study/studies
- **Samples** – The number of samples in the study/studies
- **Files** – the counts of the files

Adding Files to the Cart for Analysis

Files of interest for analysis can be selected and added to the cart as follows:

1. Select the study or PHS of interest.
2. Select all filters/metadata elements of interest.
3. The table below the cart (lower-right section of the screen) populates with the selected files.
4. Click the box for the files to be added to the cart.
5. Click **ADD SELECTED FILES**, which loads the selected files to the cart.
6. Click the cart icon on the top right of the page, which opens the cart page.



GC Cart Page

Using the Cart page, users can download a data manifest to import and use on the Velsera GCG cloud resource.

The GC Cart

HOME DATA PROGRAMS STUDIES ABOUT

Cart > Selected Files

README  AVAILABLE EXPORT OPTIONS DOWNLOAD MANIFEST 

File Name	Study Name	Accession	Participant ID	Sample ID	Study Access	File Type	Remove
SL_12629-HSYJVADXX-0-allig.cram	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	GRAM	<input type="checkbox"/>
SL_12629-HSYJVADXX-0-allig.cram.crai	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	CRAI	<input type="checkbox"/>
SL_12629-HSYJVADXX-0-chim-pe.cram	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	GRAM	<input type="checkbox"/>
SL_12629-HSYJVADXX-0-chim-pe.crai	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	CRAI	<input type="checkbox"/>
SL_12629-HSYJVADXX-0-chim-se.cram	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	GRAM	<input type="checkbox"/>
SL_12629-HSYJVADXX-0-chim-se.crai	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	CRAI	<input type="checkbox"/>
SL_12639-HSYJVADXX-0.cram	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	GRAM	<input type="checkbox"/>
SL_12639-HSYJVADXX-0.cram.crai	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	CRAI	<input type="checkbox"/>
SL_12640-HSYJVADXX-0.cram	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	GRAM	<input type="checkbox"/>
SL_12640-HSYJVADXX-0.cram.crai	Childhood Cancer Data Initiative (CCDI): Enhancement of Data Sharing in Pediatric, Adolescent and Young Adult Cancers	phs002431	00301d78915737fa100f	7219_071.9 : Brain, NOS, 7220_042.0 : Blood	Controlled	CRAI	<input type="checkbox"/>

ROWS PER PAGE: 10 1-10 OF 48

To access and analyze file click "Download Manifest" button and upload resulting file to **Seven Bridges Genomics** account

Optional: Please add subscription for the CIV file you are about to download.

1. While on the Cart page, create a description for your file in the text box on the bottom left of the screen (this is optional).
2. Users can immediately export the manifest directly to Velsera or click the **DOWNLOAD MANIFEST** button on the top-right corner of the Cart page.
3. Open the downloaded folder on your computer. The manifest will contain the elements which are selected on the cart page from the dropdown. Please note that the following metadata elements are default and will auto populate:
 - DRS URI (File ID)
 - Note that this DRS File ID will be needed when you import the manifest to the SBG-CGC
 - Name
 - Participant ID
 - Md5sumUser Comments

This manifest can be imported for analysis on SBG-CGC by following the steps below:

Exporting to SBG-CGC/Velsera

1. Create an account or log in at the following URL:
<https://www.cancergenomicscloud.org/>.
2. Create a new project or select an existing project suitable for digital access to the files listed in the file manifest.
3. Navigate to the SBG-CGC dashboard's navigation bar, choose "Files," and click the "+ Add files" button in the dropdown menu.
4. From the dropdown menu, select "GA4GH Data Repository Service (DRS)."
5. Choose "From a manifest file" and click "Browse manifest" to import the files with the DRS URI attached, generated from the GC portal. (Typically found in your downloads folder or on the desktop for easy access.)
6. Utilize the free-text search box to add relevant tags or comments associated with the batch of imported files and click "Import."
7. Check the Data Use Agreement box to confirm compliance with the data guidelines set by the SBG-CGC. Then, click submit to access your files in the created project.
8. Use free text to create tags for associating with your data.

9. DRS (Data Repository Service) identifiers will now be displayed for each file from the imported file manifest within the selected SBG-CGC project.

10. These files are accessible in the SBG-CGC Genome Browser and can be selected as inputs for downstream analysis.

How to use GraphQL interface:

The GC portal offers filtering and visualization tools to help researchers access basic information about the data stored in the GC. For more detailed insights into studies, participants, and samples, users can utilize the GraphQL interface.

GC users can access the GraphQL interface either through Jupyter or by querying it directly. Queries should be directed to the following URL:

<https://general.datacommons.cancer.gov/v1/graphql/>

Please use the 'Query the GC Database' custom query document to perform database queries.