

Contents

1	Getting Started	5
	Getting Started	5
	Introduction	5
	Accessing the GDC Data Portal	5
	Key Features	6
	Release Notes	7
2	About the Data	8
	About the Data	8
	Data Access	8
	Controlled Access Data	8
	Open Access Data	8
3	Authentication	10
	Overview	10
	Authentication via eRA Commons	10
	GDC Authentication Token	11
	Obtain a Token From GDC Data Portal	11
	Obtain a Token From GDC Data Submission Portal	12
	Logout from GDC	12
4	Portal Features	13
	Sitemap	13
	Main Menu	14
	Left Section	14
	Right Section	14
	Filtering and Searching	14
	Faceted Navigation	14
	Advanced Search	16
	Quick Search	16
	Tables	17
	Sorting	17

Columns Filtering	18
Exporting	19
Pie Charts	20
5 Projects	22
Projects Search	22
Overview	22
Available Facets	25
Graph View	25
Project Entity Page	27
6 Cases and Files	28
Cases and Files Search	28
Overview	28
Facets	29
Summary	29
Cases	30
Files	32
Advanced	33
Case Entity Page	34
File Entity Page	35
7 BAM Slicing	37
Overview	37
8 Annotations	38
Annotations Search	38
Overview	38
Facets	38
Annotation List	39
Annotation Categories and Classification	39
Annotation Entity Page	39
9 Advanced Search	41
Overview	41
Switch between ‘Advanced’ and ‘Faceted’ Search	42
Use the Advanced Search	42
Auto-complete	42
Field Auto-complete	42
Value Auto-complete	43
Setting Precedence of Operators	44

Keywords	44
AND Keyword	45
OR Keyword	45
Operators	45
“=” operator - EQUAL	46
“!=” operator - NOT EQUAL	46
“>” operator - GREATER THAN	46
“>=” operator - GREATER THAN OR EQUALS	46
“<” operator - LESS THAN	47
“<=” operator - LESS THAN OR EQUALS	47
“in” Operator	47
“not in” Operator	48
“is missing” Operator	48
“is not missing” Operator	48
Special Cases	48
Using Quotes	48
Age at Diagnosis - Unit in Days	49
Fields not Supported in the Advanced Search	49
Fields Reference	49
Files	49
Cases	50
10 Reports	52
Data Download Statistics Report	52
11 Cart and Files Download	53
Overview	53
GDC Cart	53
Download Options	53
GDC Data Transfer Tool	53
Cart Download	54
Individual Files Download	54
Access	54
12 Key Terms	55
0.2.18.3	55
GDC Data Portal Release 0.2.18.3	55
New Features and Changes	55
Bugs Fixed Since Last Release	56
Known Issues and Workarounds	56

0.2.15.4	56
GDC Data Portal Release 0.2.15-oicr4	56
New Features and Changes	56
Bugs Fixed Since Last Release	56
Known Issues and Workarounds	56
0.2.15.2	57
GDC Data Portal Release 0.2.15-oicr2	57
New Features and Changes	57
Bugs Fixed Since Last Release	57
Known Issues and Workarounds	58
0.2.15.1	58
GDC Data Portal Release 0.2.15-oicr1	58
New Features and Changes	58
Bugs Fixed Since Last Release	58
Known Issues and Workarounds	58
0.2.13	59
GDC Data Portal Release 0.2.13	59
New Features and Changes	59
Bugs Fixed Since Last Release	59
Known Issues and Workarounds	59
0.1.10	60
GDC Data Portal Release 0.1.10	60
New Features and Changes	60
Bugs Fixed Since Last Release	60
Known Issues and Workarounds	60
0.1.8	61
GDC Data Portal Release 0.1.8	61
New Features and Changes	61
Bugs Fixed Since Last Release	61
Known Issues and Workarounds	61

Chapter 1

Getting Started

Getting Started

Introduction

The National Cancer Institute (NCI) Genomic Data Commons (GDC) Data Portal User's Guide is the companion documentation to the [GDC Data Portal](#). The GDC User's Guide provides detailed information and instructions for using the [GDC Data Portal](#).

The GDC Data Portal is a robust data-driven platform that allows users to search and download cancer data sets for analysis using modern web technologies. Key GDC Data Portal features include:

- Data browsing by project, file, case, or annotation
- Visualization allowing users to perform fine-grained filtering of search results
- Data search using advanced smart search technology
- Data selection into a personalized cart
- Data download from cart or a from GDC Data Transfer Tool

Additional descriptions of key GDC Data Portal features are provided in the section that follows.

Accessing the GDC Data Portal

The GDC Data Portal is accessible using a web browser such as Chrome, Internet Explorer, and Firefox at the following URL: <https://gdc-portal.nci.nih.gov>. When navigating to the site, the GDC Data Portal main page is displayed.

The GDC Data Portal allows download of open access data without requiring a user login. To retrieve controlled access data, users must have access to the controlled access data set through dbGaP and use the Login functionality on the top-right corner of the screen.

Key Features

Data Browsing

By browsing through the GDC Data Portal, users can obtain details about specific aspects of the data, such as the associated project, files, cases and annotations (explanatory comments about the data). For example, from a project page a user can view files or cases within the project. While users are browsing through the Data Portal, they can progressively add more elements to a cart for data download.

Visualization

The GDC Data Portal allows researchers to identify data they are looking for and narrow their search using multiple visualization mechanisms such as tables, charts, and plots. Whenever possible, cross-references are enabled so that users can dive into the data directly from a plot.

Data Searching

GDC Data Portal contains two primary mechanisms for searching data: 1) searching using predefined filters (also called Facets) or 2) searching via advanced search technology and the advanced query language. When a user clicks on a Facet, the GDC smart search is automatically populated to help users get familiar with the advanced query language. At any time, when a need for a more specific query arises, users can switch to smart search and update their query with more specific filters.

Cart Facilities for Data Selection

While navigating and searching through the GDC Data Portal, researchers can add files to their cart. The cart provides detailed statistics about the files it contains to inform the user about the overall number and size of files the user may want to download.

Data Download

The GDC Data Portal provides two primary channels to allow users to download data from a cart: 1) download directly from the browser or 2) download using a dedicated Data Transfer Tool. The GDC Data Portal generates a list of desired files (or manifest) that can be easily imported into the GDC Data Transfer Tool to execute the download. Note that users can download all files that are under the GDC open access policy as well as any controlled access files that they have authorization for through dbGaP. Please visit [Obtaining Access to GDC Controlled Data](#) for information how to obtain authorization to access controlled data sets through dbGaP.

Release Notes

The GDC Data Portal is regularly being updated with new features. The Release Notes sections of the documentation contains details about new features, bug fixes and known issues.

Chapter 2

About the Data

About the Data

The GDC provides access to data from several cancer research programs including The Cancer Genome Atlas (TCGA), Therapeutically Applicable Research to Generate Effective Treatments (TARGET), and the Cancer Genome Characterization Initiative (CGCI).

For all new programs and projects data sets added to GDC, the GDC Data Portal provides access to these submitted data sets, which comply with GDC's data standards for biospecimen, clinical, and molecular data.

For additional information on current GDC supported programs, the following resources are available:

- [TCGA Data Primer](#)
- [TARGET Data](#)
- [CGCI Data](#)

Data Access

Controlled Access Data

The GDC provides access to open and controlled data. Example of controlled access data include: genomic sequence, germline variants, SNP6 genotype data, and possibly certain clinical data elements depending on the program.

Access to controlled data is granted by program-specific Data Access Committees. Anyone who wishes to download controlled access data from the GDC must follow the same set of steps required by NCBI for access to this data in dbGaP. That is, an eRA Commons account must be obtained, and then access must be approved by the appropriate Data Access Committee. Instructions for how to obtain authorization for controlled access data can be found at the following link: https://dbgap.ncbi.nlm.nih.gov/aa/dbgap_request_process.pdf

To download Controlled Access data, users must login to the GDC Data Portal using an eRA commons username and password.

Open Access Data

The GDC Data Portal provides access to data that is publicly available. Open access data includes all Level 3 data (somatic mutations, gene expression, protein expression), as well as most clinical and all biospecimen data elements.

Users can browse information in the GDC Portal without having to login. Anything that can be seen without logging in is Open Access data or metadata.















Logging In

While not required for public data access, a user may login using the 'login' button at the upper right side of GDC Data Portal opening screen.

More details about Authentication and Authorization can be found in a dedicated section of the documentation.

GDC Data Portal Display of Controlled and Open Access Data

All metadata displayed in the GDC Data Portal is open access. When data files are displayed, there will be a column in the display that indicates whether the file is open or controlled access.

Summary Cases (11,484) Files (558,853)					
Files					
Showing 1 - 20 of 558,853 files					
 	Access	File Name	Case UUIDs	Project	
 	 Controlled	000aa811c15656604161e8...	1	TCGA-SARC	
 	 Controlled	0017ba4c33a07ba807b291...	1	TCGA-BRCA	
 	 Controlled	00286ada95aed3619130f0...	1	TCGA-UCEC	
 	 Controlled	003860a34c9b244a5d8435...	1	TCGA-BRCA	

The 'lock' icon represents whether a file is controlled or open access. It does not indicate whether a particular user has access to that file. As mentioned above, if the 'My Projects' option is checked, then a user will only see files that they have access to.

Chapter 3

Authentication

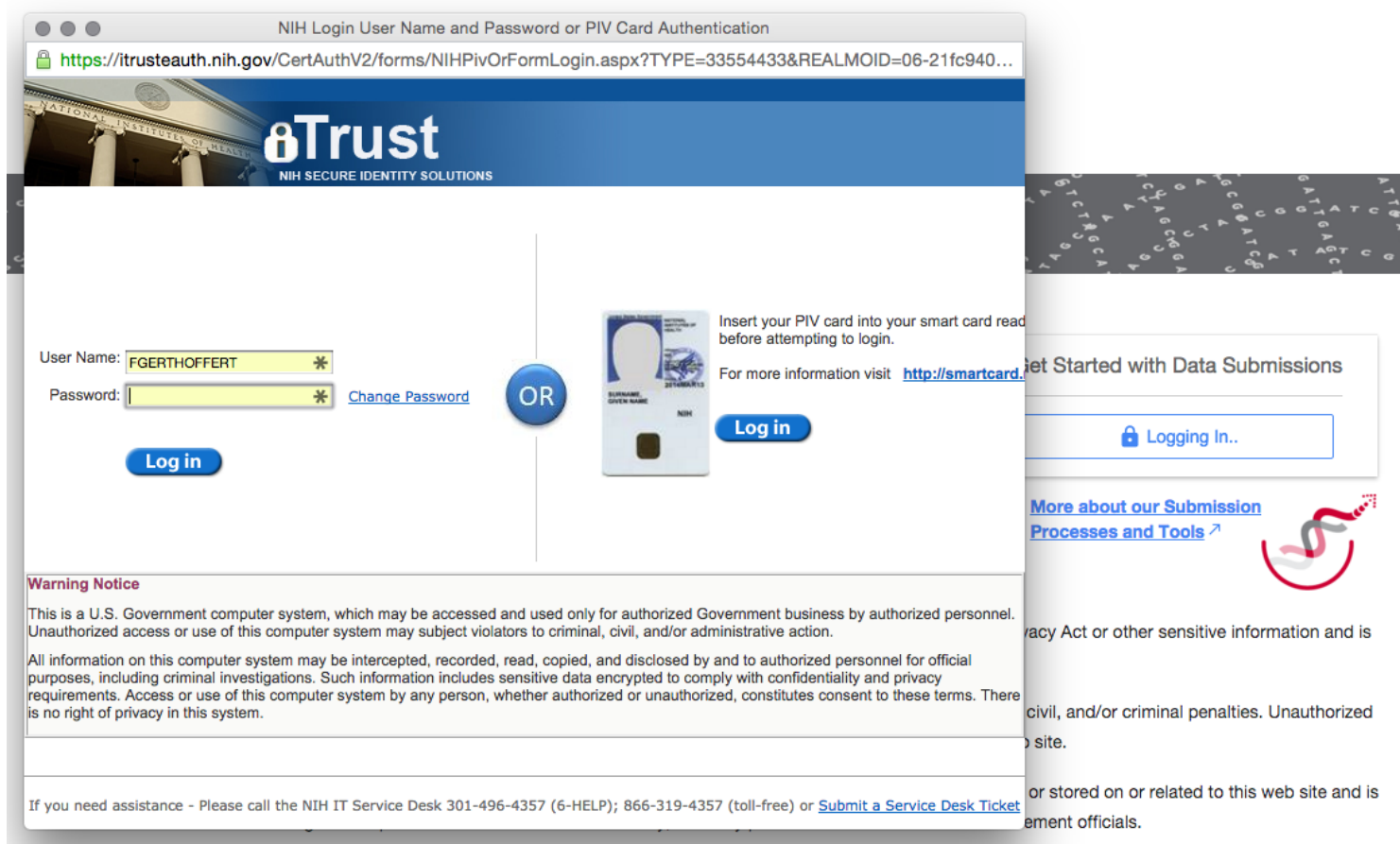
Overview

Accessing the GDC Data Submission Portal requires eRA Commons credentials with appropriate dbGap authorization. See [Obtaining Access to Submit Data](#) for more details.

Permissions are attached to each user account. Access can be granted to one or more projects. Those permissions will apply throughout the GDC Data Portal and GDC Data Submission Portal and will define how data is being displayed.

Authentication via eRA Commons

By Clicking on “Login”, users will be invited to authentication through eRA Commons. If successful, the user will be redirected to the GDC Data Portal or GDC Data Submission Portal.



GDC Authentication Token

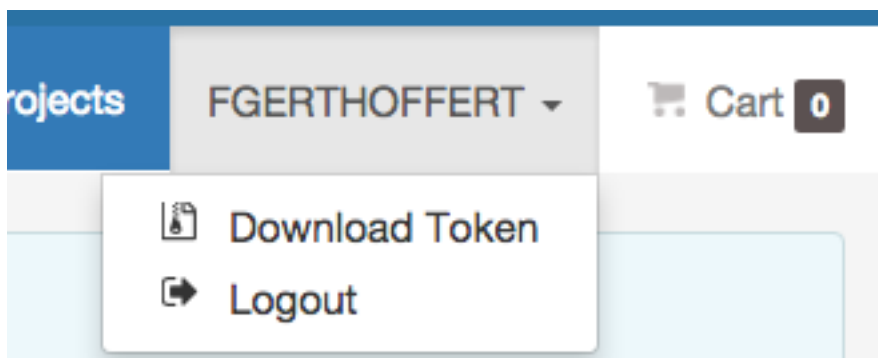
A Token is required to upload data to GDC, as well as to download controlled-access files, using the GDC Data Transfer Tool or the GDC Application Programming Interface.

The GDC Data Transfer Tool is optimized for large transfers with multi-part upload and integrity checking, making it the most efficient tool for molecular data submission.

Tokens are strings of characters provided with every GDC call requiring authentication. A file containing a token can be obtained through GDC Data Submission Portal once your NIH eRA Commons account has been established. Your GDC account will be automatically established upon first login and you will see a drop down menu next to your user name.

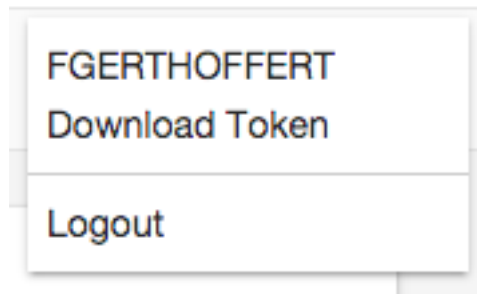
Obtain a Token From GDC Data Portal

After successful authentication, users can obtain their Token by clicking on their username on the top right corner of the screen.



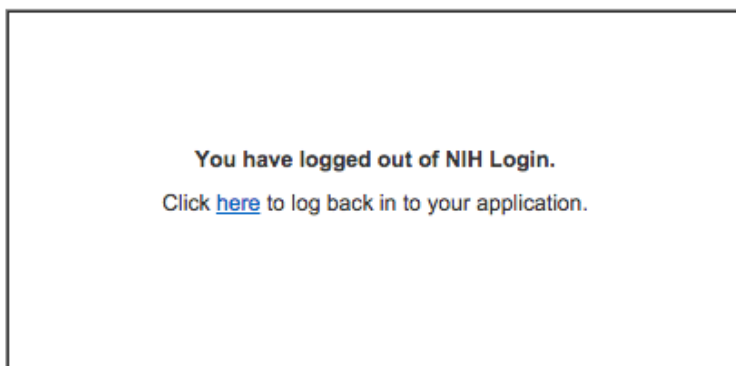
Obtain a Token From GDC Data Submission Portal

After successful authentication, users can obtain their Token by clicking on their username on the top right corner of the screen.



Logout from GDC

By clicking on logout, GDC will be redirected to an eRA Commons page, displaying the logout status.



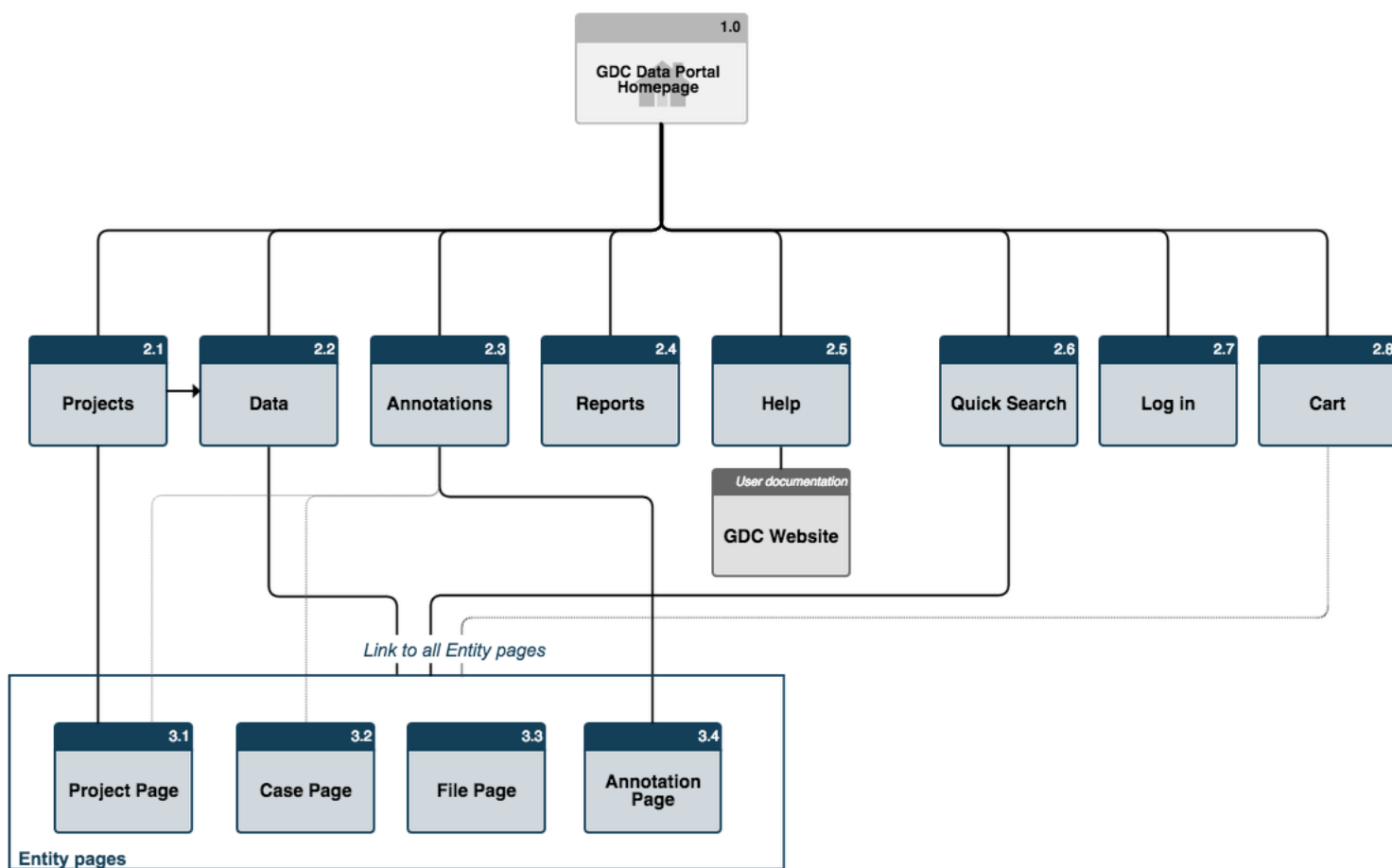
If you need assistance - Please call the NIH IT Service Desk call 301-496-4357 (local), 866-319-4357 (toll-free), or 301-496-8294 (TTY). or [Submit NIH IT Service Desk Ticket](#)



Chapter 4

Portal Features

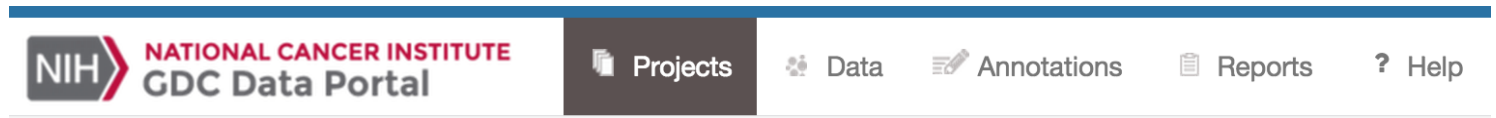
Sitemap



The portal is composed by 8 major sections available to the user through a toolbar on the upper level of the GDC Data Portal.

Main Menu

Left Section

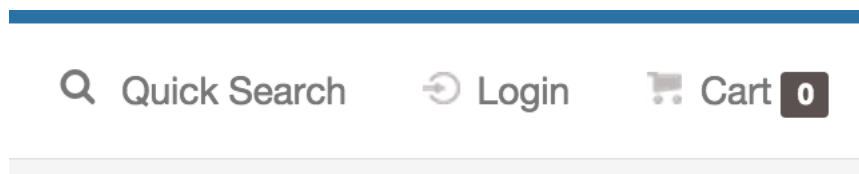


The left portion of this toolbar provides access to the following elements:

- **Projects:** High-level overview of all data
- **Data:** Search through all cases and files
- **Annotations:** Search through all annotations
- **Reports:** Reports about data available
- **Help:** A link to this user guide

Each *'Projects'*, *'Data'*, and *'Annotations'* view enables the user to set filters from within the view to either browse the data or select it for export or download. It provides different and independent ways of organizing GDC data and metadata.

Right Section



The right portion of this toolbar provides access to the following elements:

- **Quick Search:** Quickly search for more information about a specific entity
- **Login:** Log into the GDC Data Portal through eRA Commons
- **Cart:** Access to user populated cart mechanism

Filtering and Searching

The GDC Data Portal provide three different means to search and navigate throughout available data:

- **Faceted Navigation:** By clicking on relevant items on the left portion of *'Projects'* & *'Data'* views, a user can filter down displayed elements
- **Advanced Search:** Provides access to advanced and complex queries using a purpose-built query language.
- **Quick Search:** Provide, from any place on the GDC Data Portal, a quick access to specific entities

Faceted Navigation

Available on the left side of *'Projects'* & *'Data'* views, *'Faceted Navigation'* enable filtering by selecting relevant data. Facets are attributes of the data that are being searched. Faceted Navigation filters the overall data set based on the set of Facets chosen.

'Faceted Navigation' is composed by three elements:

- **Query facets:** used to select relevant data
- **Query Field:** showing filters currently being applied
- **Results:** showing filtered results corresponding to the query

▼ Primary Site

- ☐ Blood 6
- ☐ Kidney 6
- ☐ Brain 2
- ☐ Colorectal 2
- ☐ Lung 2

[25 More...](#)

Note: In facets, the end of each filter line contains a number. This number corresponds to the number of entities corresponding to this filter.

By clicking on an element, a filter is automatically applied. If the user clicks on multiple elements, an **OR** filter will be applied and the query will be displayed at the top of the page.

The count (in grey cell) next to each Facet, indicates the number of results corresponding to the field.

Primary Site **IN** (**Kidney** , **Brain**) **AND** **Data Type** **IN** (**Clinical** , **Gene Expression**)

By clicking on a query term will automatically remove the corresponding filter.

The above query could be translated to:

Return all projects where primary site is Kidney OR Brain and containing Data Type Clinical OR Gene Expression

The resulting set is displayed below the query.

Table	Graph
-------	-------

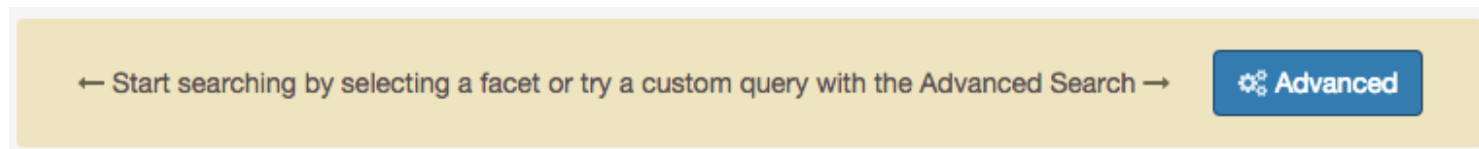
Projects (6)

↓^A_Z
≡
⌘

ID	Disease Type	Primary Site	Program	Cases	Available Cases				
					Clinical	Array	Seq	SNV	CNV
TCGA-GBM	Glioblastoma Multif...	Brain	TCGA	617	617	613	442	602	602
TCGA-KIRC	Kidney Renal Clear ...	Kidney	TCGA	537	537	536	535	534	534
TCGA-LGG	Brain Lower Grade ...	Brain	TCGA	516	516	516	516	516	515
TCGA-KIRP	Kidney Renal Papilla...	Kidney	TCGA	291	291	291	291	290	290
TARGET-WT	High-Risk Wilms Tu...	Kidney	TARGET	128	128	126	128	128	124
TCGA-KICH	Kidney Chromophobe	Kidney	TCGA	113	113	66	66	66	66

Advanced Search

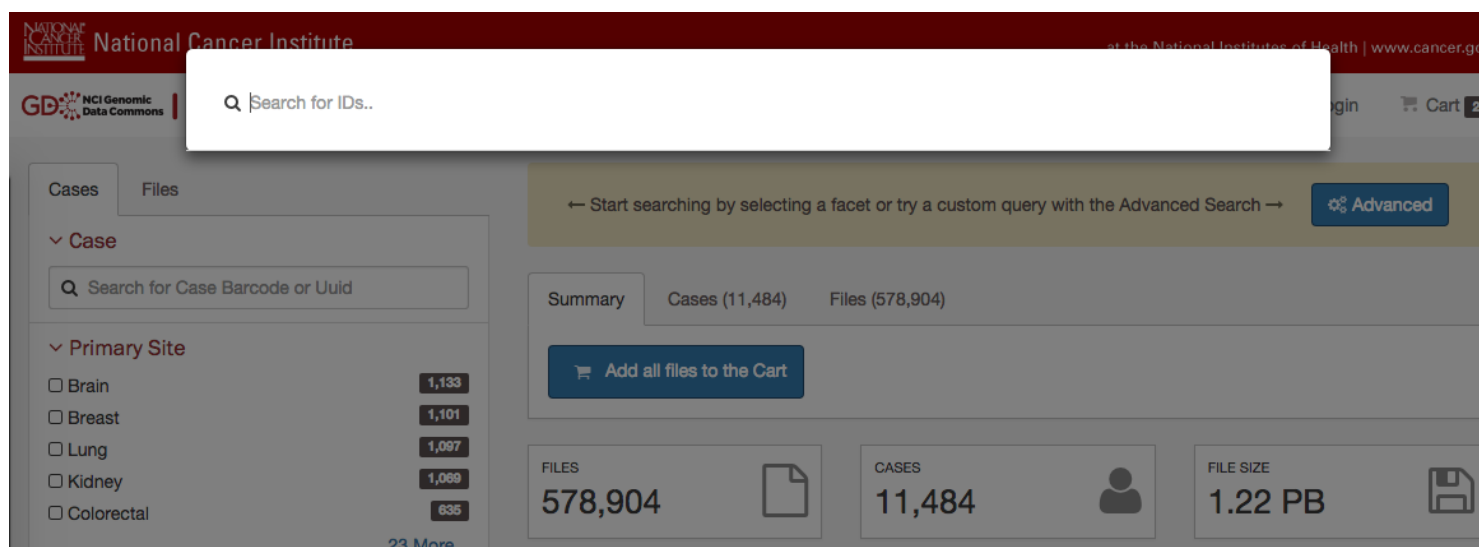
Only available in the ‘Data’ view, the Advanced Search page offers complex query building capabilities to identify specific set of cases and files.



The Advanced Search section of the documentation provides more details about this feature.

Quick Search

The GDC Data Portal is equipped with a Quick search feature available either by clicking on the magnifier in the right section of the toolbar or by pressing CTRL SPACE on a keyboard.



Search results are automatically displayed while the user is typing.

Q 0019|

001935b9-684d-4e1a-818f-66450e46726d

00193eac-8fb6-4a47-859d-85e8f719ff09

001944e5-af34-4061-9c09-bb9ea346f6fd

0019771f-ebe0-41c3-8ef7-5ccae5bf246b

001905fd-e110-40ea-83bd-5bd506978c4c

File UUID	001935b9-684d-4e1a-818f-66450e46726d
File Name	bcgsc.ca.TCGA-20-1683-01A-01R-1566-13.hg19.bcgsc.c a.1.7.0.sv.vcf
Data Type	Structural rearrangement
Size	414 KB

Pressing enter or clicking on the result will open the corresponding entity page.

Tables

Most tables available throughout the GDC Data portal are equipped with export, sorting and filtering capabilities.

Table

Graph

Projects (46)

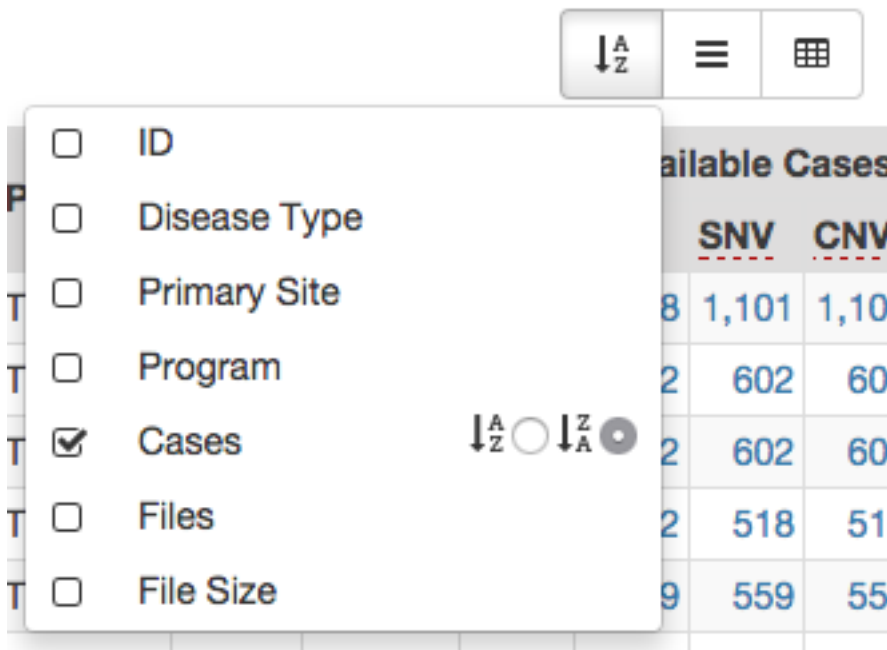
↓^A_Z

≡

ID	Disease Type	Primary Site	Program	Cases	Available Cases				
					Clinical	Array	Seq	SNV	CNV
TCGA-BRCA	Breast Invasive Carc...	Breast	TCGA	1,101	1,101	1,101	1,098	1,101	1,101
TCGA-GBM	Glioblastoma Multif...	Brain	TCGA	617	617	613	442	602	602

Sorting

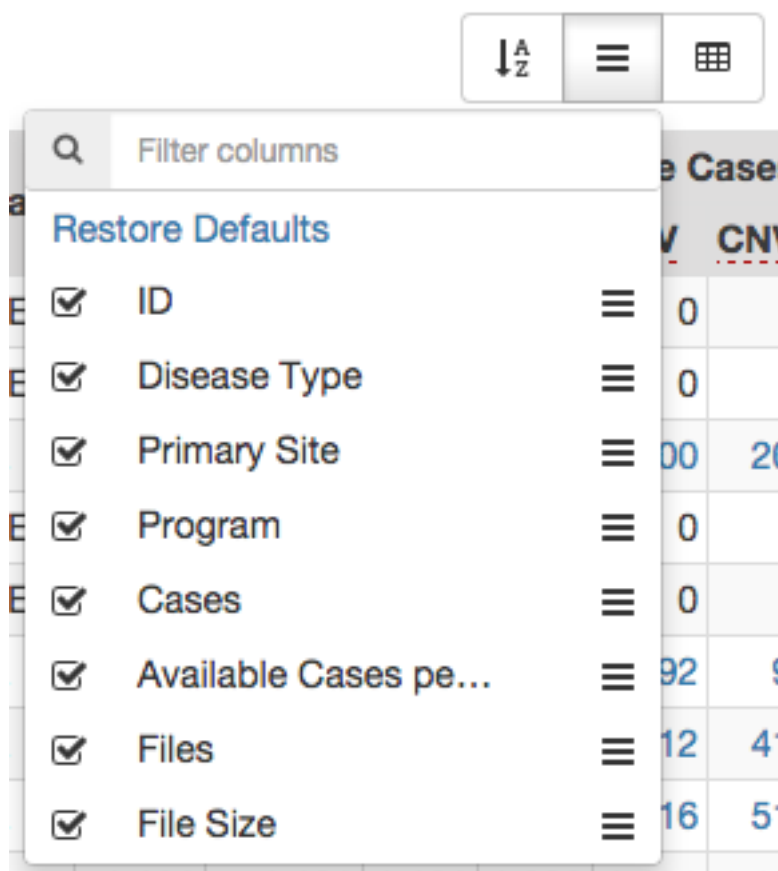
Users can sort content by selecting the corresponding column to apply sorting to.



In situations where multiple columns are selected for sorting, the system will take columns in order and progressively apply sorting. If a column with unique values (such as ID) is selected as the first column for sorting, sorting will not be available for other columns selected (since sorting is first applied on unique values subsorting is not possible).

Columns Filtering

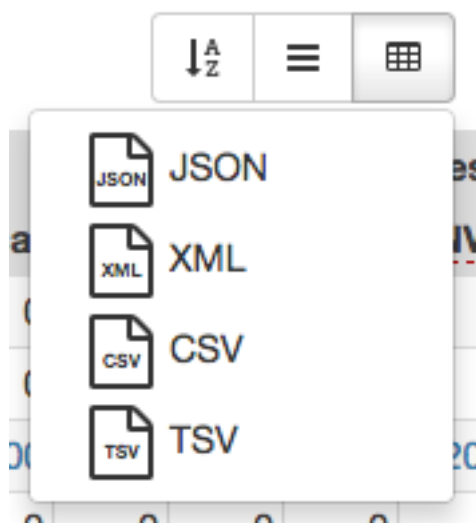
Users can select which column they want to display and in which order, preferences will be maintained throughout user visits. By default not all columns might be displayed in the table, this options provides access to advanced values.



Exporting

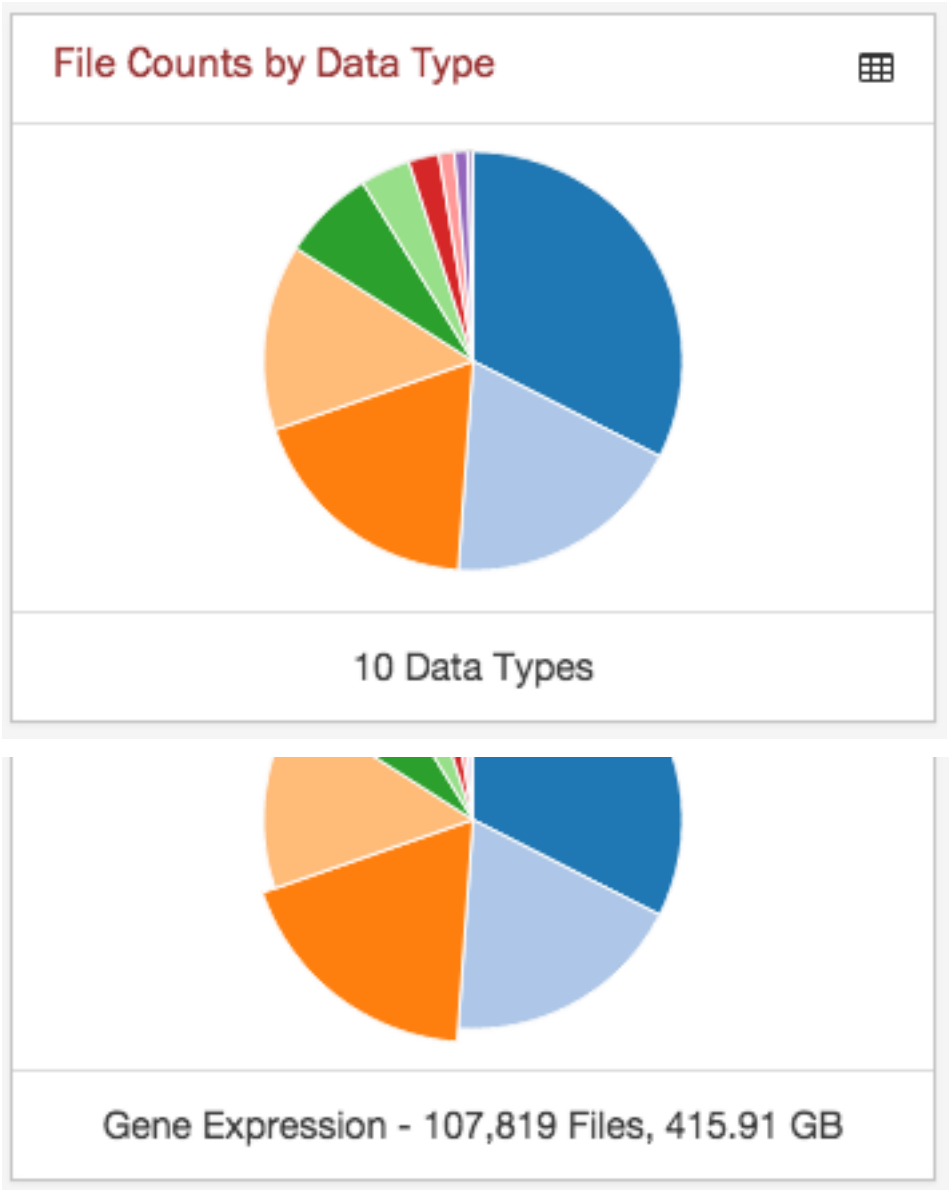
Tables can be exported in four different formats:

- JSON
- XML
- CSV
- TSV



Pie Charts

Most Pie charts in the GDC Data Portal are interactive. Mousing over a slice of the pie chart display detailed about the slice. Users are also offered the option to switch to a table view for more details.



The pie chart can also be converted to a table by clicking on an small table icon available on the top right corner of the widget.

File Counts by Data Type



	Files	File Size
Copy Number Variat...	188,851	3.53 TB
Raw Microarray Data	108,249	4.55 TB
Gene Expression	107,819	415.91 GB
Raw Sequencing Data	83,315	1.21 PB
Simple Nucleotide V...	41,516	2.13 TB
Clinical	22,411	1.38 GB
DNA Methylation	13,698	770.21 GB
Other	6,898	354 MB
Protein Expression	5,745	28 MB
Structural Rearrang...	2,501	782 MB

Chapter 5

Projects

Projects Search

Overview

The Projects view provides an overview of all the data contained in the GDC Data Portal, organized by Project. Administratively, a *Project* is a specific effort undertaken as part of an overarching Program. Many projects were set up to study one tumor type, so project names and disease names can often be the same, as is the case for TCGA data, but this is not always the case.

When accessing the Projects view, the initial screen will display a table summarizing the information about the cases (i.e. sample donors) and data types for which data is available for each project. This table displays for each project, the number of cases, number of files of each data type contained in the GDC for that project, and total file size of all the files in the project. With no Facets selected, this view includes all GDC Data Portal data.

This view focuses on the number of cases available at GDC within each program and project. Each case will have multiple types of data associated with it. Selections in the facets allows filtering of the number of cases in each project to those that have associated data files relevant to any of the selected data types or selected experimental strategies.

NIH

NATIONAL CANCER INSTITUTE

GDC Data Portal

Projects

Data

Annotations

Reports

Help

Quick Search

Login

Cart

Project

Search for Project Name or ID

Primary Site

☐ Blood

6

☐ Kidney

6

☐ Brain

2

☐ Colorectal

2

☐ Lung

2

25 More...

Cancer Program

☐ TCGA

37

☐ TARGET

9

Disease Type

☐ Acute Lymphoblastic Leukemia

2

☐ Acute Myeloid Leukemia

2

☐ Not Applicable

2

☐ Acute Myeloid Leukemia Induction F...

1

☐ Adrenocortical Carcinoma

1

38 More...

Data Type

☐ Raw Sequencing Data

37

☐ Clinical

35

☐ Copy Number Variation

35

☐ DNA Methylation

35

☐ Gene Expression

35

5 More...

Experimental Strategy

☐ RNA-Seq

37

☐ WXS

37

Start searching by selecting a facet

Table

Graph

Projects (46)

12

ID	Disease Type	Primary Site	Program	Cases	Available Cases per Data Type												Files	File Size
					Clinical	Array	Seq	SNV	CNV	SV	Exp	PExp	Meth	Other				
TARGET-NBL	Neuroblastoma	Nervous System	TARGET	1,178	0	0	801	0	0	0	0	0	0	0	1,732	115.78 TB		
TCGA-BRCA	Breast Invasive Carci...	Breast	TCGA	1,101	1,101	1,101	1,098	1,101	1,101	0	1,101	410	1,084	0	52,601	70.66 TB		
TARGET-AML	Acute Myeloid Leuke...	Blood	TARGET	894	0	0	853	0	0	0	0	0	0	0	2,836	204.15 TB		
TARGET-WT	High-Risk Wilms Tumor	Kidney	TARGET	665	128	126	652	125	124	62	63	0	125	0	3,316	75.20 TB		
TCGA-GBM	Glioblastoma Multifor...	Brain	TCGA	617	617	613	442	602	602	0	593	214	599	0	32,966	74.08 TB		
TCGA-OV	Ovarian Serous Cysta...	Ovary	TCGA	608	608	605	582	602	603	423	593	412	602	0	46,110	106.42 TB		
TCGA-LUAD	Lung Adenocarcinoma	Lung	TCGA	586	586	581	582	518	518	0	520	181	580	0	25,900	53.98 TB		
TCGA-UCEC	Uterine Corpus Endo...	Uterus	TCGA	560	560	559	559	559	559	0	559	200	559	535	27,851	34.42 TB		
TCGA-KIRC	Kidney Renal Clear C...	Kidney	TCGA	537	537	536	535	534	534	0	535	0	535	0	27,086	62.31 TB		
TCGA-HNSC	Head and Neck Squa...	Head and Neck	TCGA	528	528	528	528	528	526	0	528	212	528	0	25,900	57.19 TB		
TCGA-LGG	Brain Lower Grade Gl...	Brain	TCGA	516	516	516	516	516	515	0	516	430	516	0	24,679	40.61 TB		
TCGA-LUSC	Lung Squamous Cell ...	Lung	TCGA	511	511	511	504	506	506	0	511	195	505	0	26,584	53.87 TB		
TCGA-THCA	Thyroid Carcinoma	Thyroid	TCGA	507	507	507	507	507	507	0	507	372	507	0	25,184	50.81 TB		
TCGA-PRAD	Prostate Adenocarcin...	Prostate	TCGA	500	500	500	498	500	498	0	498	352	498	0	23,884	35.04 TB		
TCGA-STAD	Stomach Adenocarci...	Stomach	TCGA	478	478	478	443	478	478	288	461	392	478	460	25,215	56.31 TB		
TCGA-SKCM	Skin Cutaneous Mela...	Skin	TCGA	471	471	471	470	471	471	0	470	204	471	0	20,753	41.43 TB		
TCGA-COAD	Colon Adenocarcinoma	Colorectal	TCGA	463	463	461	460	461	461	0	461	332	459	458	26,750	65.94 TB		
TCGA-BLCA	Bladder Urothelial Ca...	Bladder	TCGA	412	412	412	412	412	412	0	412	127	412	0	19,886	29.96 TB		
TCGA-LIHC	Liver Hepatocellular ...	Liver	TCGA	377	377	377	377	377	377	0	376	0	377	0	15,983	64.83 TB		
TCGA-CESC	Cervical Squamous C...	Cervix	TCGA	308	308	308	307	305	305	0	308	173	307	0	12,865	16.19 TB		
TCGA-KIRP	Kidney Renal Papillar...	Kidney	TCGA	291	291	291	291	290	290	0	291	215	291	0	14,162	45.88 TB		
TCGA-SARC	Sarcoma	Mesenchymal	TCGA	261	261	261	261	261	261	0	261	223	261	0	13,035	16.22 TB		
TCGA-LAML	Acute Myeloid Leuke...	Blood	TCGA	200	200	200	200	200	200	179	198	0	194	0	13,000	36.59 TB		

Table

Graph

Projects (37)

ID	Disease Type	Primary Site	Program	Participants	Clinical
TCGA-BRCA	Breast Invasive Carcinoma	Breast	TCGA	1,101	1,101
TCGA-GBM	Glioblastoma Multiforme	Brain	TCGA	617	617
TCGA-OV	Ovarian Serous Cystadenocarcinoma	Ovary	TCGA	608	608
TCGA-LUAD	Lung Adenocarcinoma	Lung	TCGA	586	586
TCGA-UCEC	Uterine Corpus Endometrial Carcinoma	Uterus	TCGA	560	560
TCGA-KIRC	Kidney Renal Clear Cell Carcinoma	Kidney	TCGA	537	537
TCGA-HNSC	Head and Neck Squamous Cell Carcinoma	Head and Neck	TCGA	528	528
TCGA-LGG	Brain Lower Grade Glioma	Brain	TCGA	516	516
TCGA-LUSC	Lung Squamous Cell Carcinoma	Lung	TCGA	511	511
TCGA-THCA	Thyroid Carcinoma	Thyroid	TCGA	507	507
TCGA-PRAD	Prostate Adenocarcinoma	Prostate	TCGA	500	500
TCGA-STAD	Stomach Adenocarcinoma	Stomach	TCGA	478	478
TCGA-SKCM	Skin Cutaneous Melanoma	Skin	TCGA	471	471

Export Table ▼

Participants	Available Participants per Data Type										Files	File Size
	Clinical	Array	Seq	SNV	CNV	SV	Exp	PExp	Meth	Other		
1,101	1,101	1,101	1,098	1,094	1,101	0	1,101	410	1,084	0	52,399	70.69 TB
617	617	613	442	602	602	0	593	214	599	0	31,986	73.84 TB
608	608	603	582	602	603	0	593	0	602	0	38,651	102.84 TB
586	586	581	582	511	518	0	520	181	580	0	25,405	53.68 TB
560	560	559	559	559	559	0	559	200	559	501	27,742	34.44 TB
537	537	536	535	528	534	0	535	0	535	0	26,616	62.06 TB
528	528	528	528	526	526	0	528	212	528	0	24,878	57.66 TB
516	516	516	516	514	515	0	516	258	516	0	22,186	29.60 TB
511	511	511	504	506	506	0	511	195	505	0	26,430	53.44 TB
507	507	507	507	507	507	0	507	372	507	0	24,284	50.62 TB
500	500	500	498	500	498	0	498	164	498	0	23,088	34.93 TB
478	478	478	443	478	478	36	477	299	478	330	22,783	53.84 TB

Clicking on any cell in the Project View Table will display filtered results in the ‘Data’ view described in this documentation.

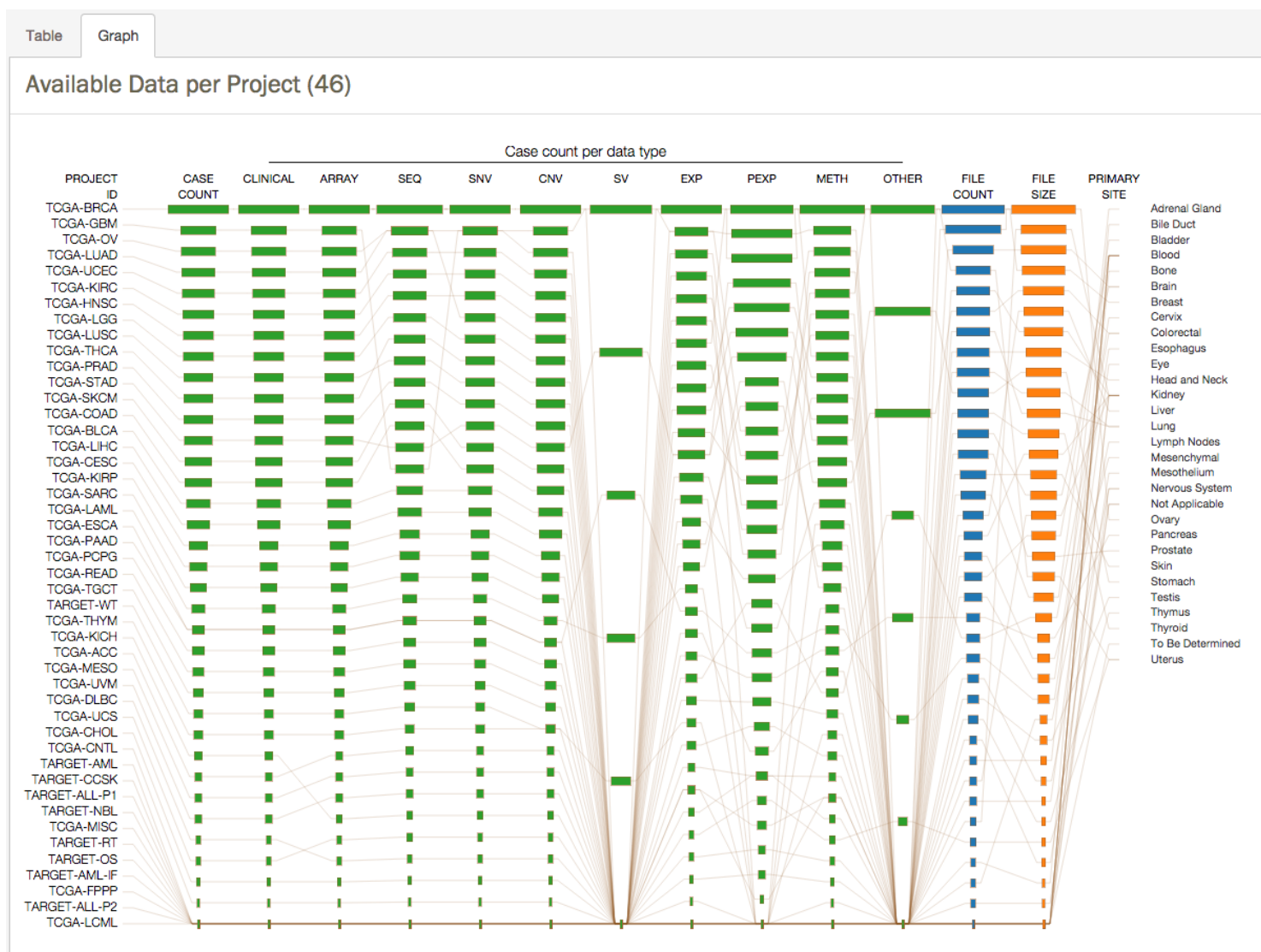
Available Facets

As mentioned previously in this documentation, the projects view offer the ability to filter results using facets. The following facets are available in the projects page:

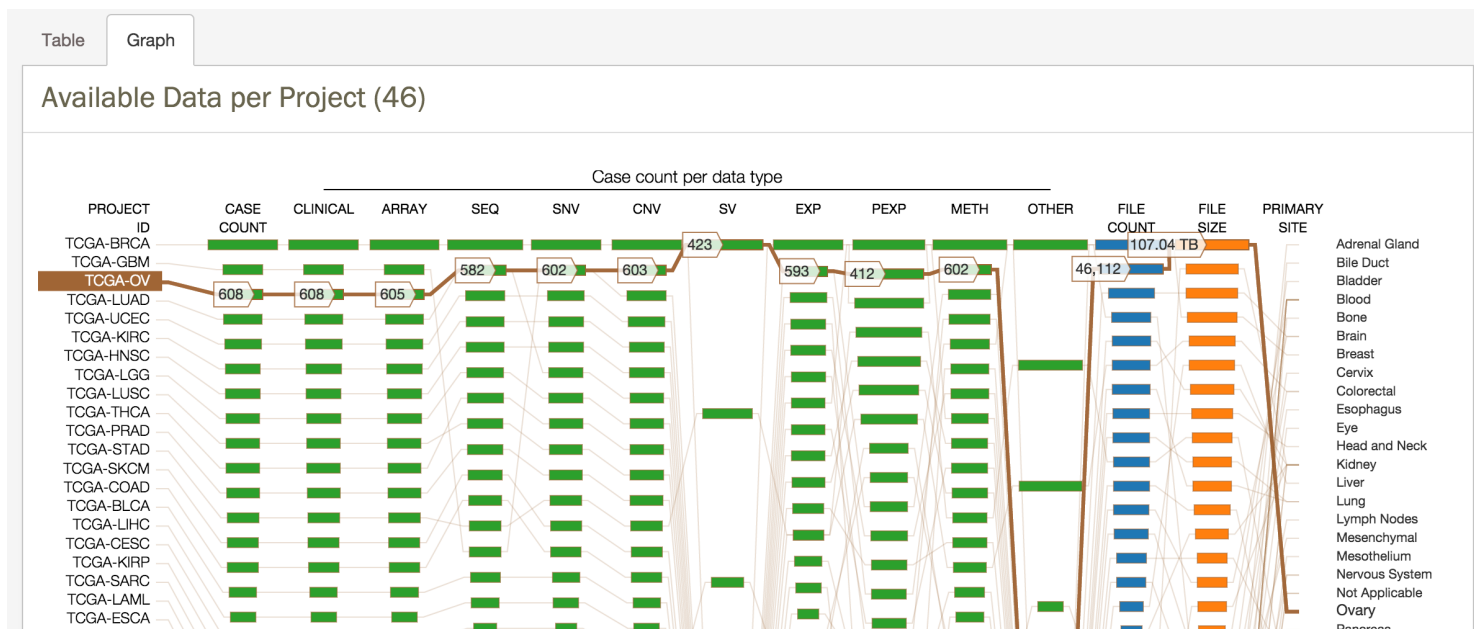
- **Project:** Projects are described above.
- **Primary Site:** Originating or primary anatomic site of the cancer under investigation or review.
- **Cancer Program:** Programs are overarching activities meant to fulfill a broad scientific objective. As mentioned above, Projects fulfill more specific requirements within a Program.
- **Disease Type:** type of cancer studied in the project.
- **Data Type:** Kind of data contained in the file. Note that each of these data types may have subtypes, and that the numbers given in the table reflect the numbers of latest files for all subtypes of that data type. Data Subtypes can be explored in Facets contained in the ‘Files’ tab in the ‘Data’ view described below.
 - *Clinical* – Clinical Data about the case, as well as Biospecimen data about the sample and aliquots
 - *Array* – Gene Expression Array Data
 - *Seq* – Raw Sequencing Data
 - *SNV* – Single Nucleotide Variation
 - *CNV* – Copy Number Variation
 - *SV* – Structural Rearrangement
 - *Exp* – Raw Microarray Data
 - *PExp* – Protein Expression
 - *Meth* – Methylation
 - *Other* – a catch-all for any other data associated with a particular sample
- **Experimental Strategy:** This describes NGS sequencing strategies or microarray technological platforms/array types and other experimental assays.

Graph View

Clicking the ‘Graph’ tab will display an interactive view of the data currently being displayed according to selected filters. The top number in each column representing the highest count of that type.



When a user mouse-over an element of the graph, the whole line corresponding to the project is automatically highlighted



This graph has been configured to be clickable, on the above example, clicking on the 'SV' element will automatically redirect the

user to the Data view with the following filter applied:

↶ Clear

Data Type

IS

Structural Rearrangement

AND

Project Id

IS

TCGA-OV

⚙️ Advanced

Project Entity Page

Each project has its own entity page aimed at providing an overview of all cases, files and annotations available for the selected project.

TCGA-BRCA

Summary

Id	TCGA-BRCA
Project Name	Breast Invasive Carcinoma
Disease Type	Breast Invasive Carcinoma
Primary Site	Breast
Program	TCGA
Cases	1,101
Files	52,607
Annotations	1,452

Experimental Strategy

Analysis	Cases	Files
Methylation Array	1,084	4,904
Gene Expression Array	532	1,791
Rna-Seq	1,097	12,468
Protein Expression Array	410	1,104
Mirna-Seq	1,084	4,643
Wxs	1,084	2,247
Total Rna-Seq	12	222
Validation	139	321
Bisulfite-Seq	5	24
Genotyping Array	1,101	22,430
Wgs	118	264

Available Data

Data Type	Cases	Files
Clinical	1,101	2,189
Raw Microarray Data	1,101	7,836
Raw Sequencing Data	1,098	7,552
Simple Nucleotide Variation	1,101	2,270
Copy Number Variation	1,101	17,909
Structural Rearrangement	0	0
Gene Expression	1,101	13,203
Protein Expression	410	410
Dna Methylation	1,084	1,238
Other	0	0

Clicking on various numbers will link to the corresponding search page (data or annotations) with filters already applied. For example, if the cases link in the ‘Clinical’ row in the ‘Available Data’ Table is chosen, the user will see a case view with the appropriate filters already applied.

Chapter 6

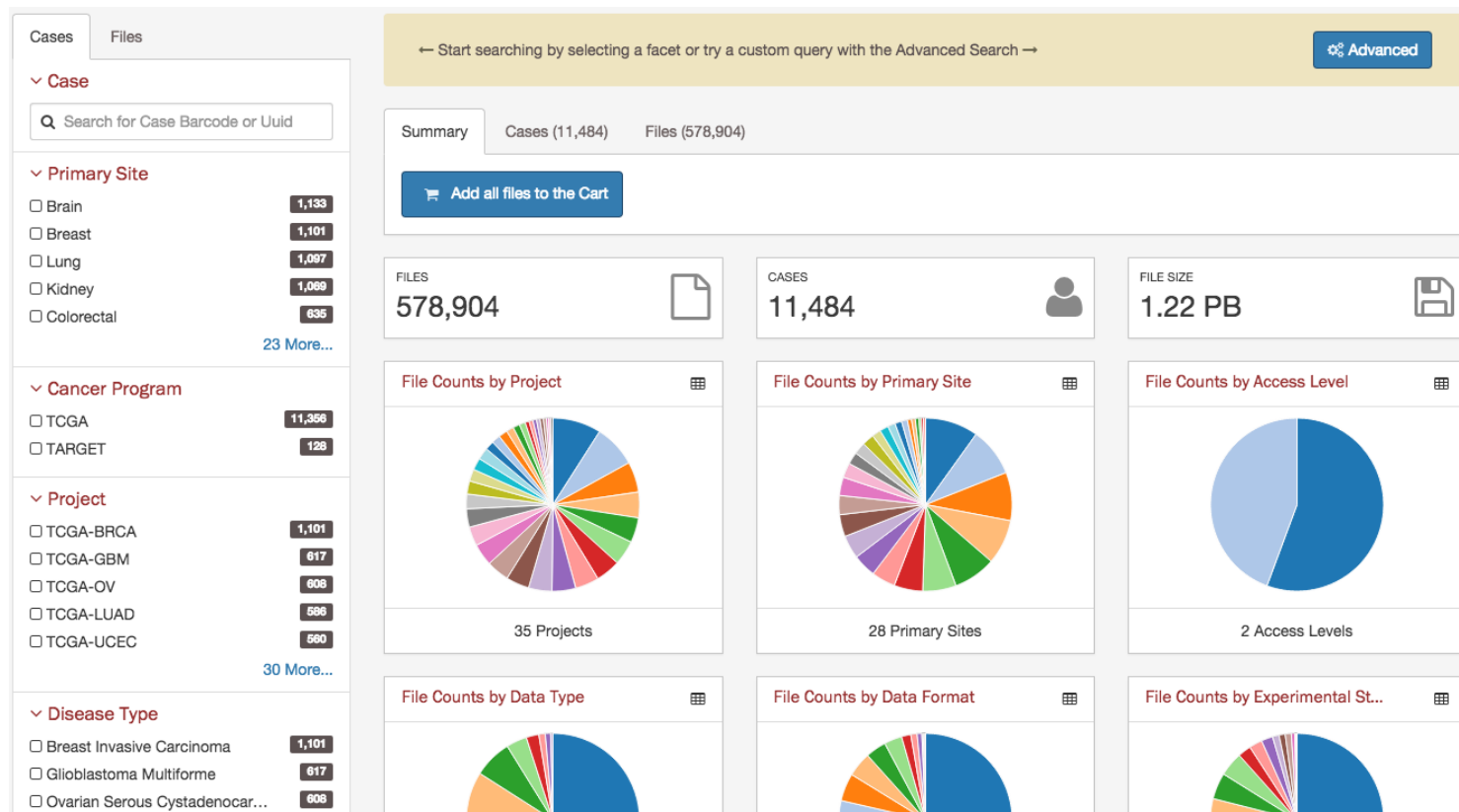
Cases and Files

Cases and Files Search

The data view is the entry point to data searching on the GDC Data Portal, it provides access to all main features of the portal

Overview

The Data view is the main section used on the GDC Data Portal to search for data.



This view focuses on the number of data files available at GDC. Selecting items from the Cases and Files facet will reduce the numbers of files displayed to those files which have all the selected properties. Note that in this view, when a choice is selected (for example, Copy Number Variation), other facets will change to hide properties that are incompatible with the selected choice (for example, when Copy Number Variation is selected in the Files: Data Type facet, only data subtypes relevant to CNV are visible in the Files: Data Subtype facet).

Facets

Two types of facets are available, '*Cases*' and '*Files*' facets.

By default, the following **Cases** facets are available:

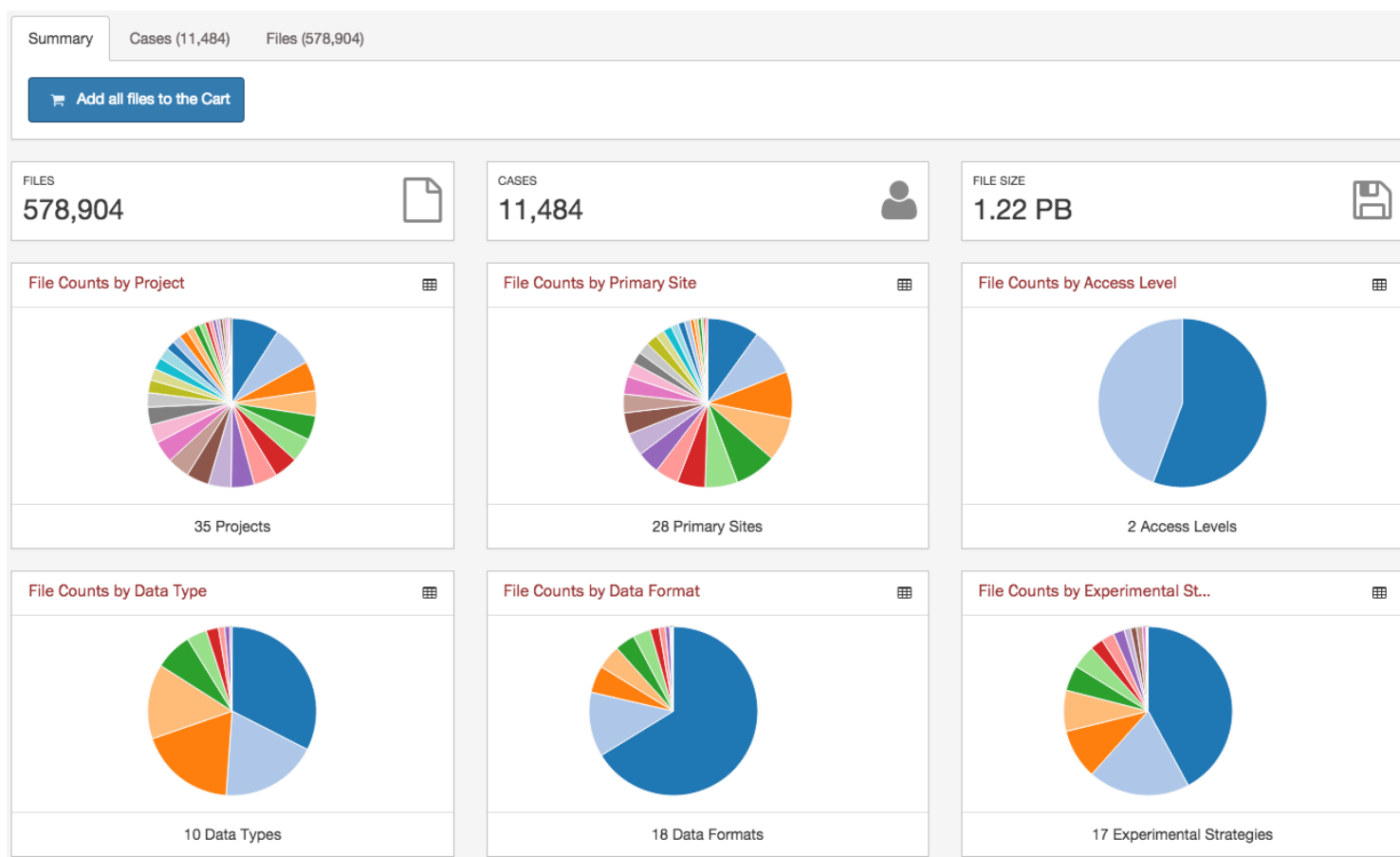
- **Case:** Search for cases using barcode or UUID.
- **Primary Site:** Originating or primary anatomic site of the cancer under investigation or review.
- **Cancer Program:** Programs are overarching activities meant to fulfill a broad scientific objective. As mentioned above, Projects fulfill more specific requirements within a Program.
- **Project:** See project section of the documentation.
- **Disease Type:** Type of cancer studied in the project.
- **Gender:** Female or Male.
- **Age at diagnosis:** Patient age at the primary diagnosis.
- **Vital Status:** Indicate whether the patient was living or deceased at the date of last contact (Alive or Dead).
- **Days to death:** Number of days between primary diagnosis and death of the patient.
- **Race:** Each of the major divisions of humankind, having distinct physical characteristics.
- **Ethnicity:** Ethnicity is a socially defined category of people who identify with each other based on common ancestral, social, cultural or national experience.

By default, the following **Files** facets are available:

- **File:** Search for files using filename or ID
- **Data Type:** Kind of data contained in the file. Note that each of these data types may have subtypes, and that the numbers given in the table reflect the numbers of latest files for all subtypes of that data type.
- **Experimental Strategy:** This describes NGS sequencing strategies or microarray technological platforms/array types and other experimental assays.
- **Data Format:** Format of the data.
- **File Origin:** Indicate whether file is migrated from legacy data, submitted (or produced by harmonization) in active projects.
- **Platform:** Technological platform from which experimental data was produced.
- **Access Level:** Indicate whether data is open or controlled access.
- **Data Submitter:** Research center from where data was produced.
- **File Status:** Indicate status of data file, possible values: live, redacted etc.
- **Tags:** Short keyword or phrase assigned to data file to describe any perspective of the data file to facilitate data search.

Summary

The summary tab displays a high-level view of the data currently being filtered.



On top of the page a button is available to add all filtered files to cart.

Multiple pie charts are available to provide a visual and interactive representation of data available on the GDC Data Portal. Clicking on a specific slice of a table filter add the corresponding slice to currently applied filters.

Cases

The Cases tab provides a list of all cases available on the GDC Data Portal. As with other sections of the portal, results can be filtered down via facets. Looking at the query field on top of the page is the best way to easily identify if a filter is currently being applied.

Summary

Cases (11,484)

Files (578,904)

Cases

Showing 1 - 20 of 11,484 cases

↓^A₂

≡

☰

Cart	Case ID	Project	Primary Site	Gender	Files	Available Files per Data Type											Annotations
						Clinical	Array	Seq	SNV	CNV	SV	Exp	PExp	Meth	Other		
<div>🛒</div>	0004d251-3f70-4395-...	TCGA-LIHC	Liver	Male	39	2	5	5	2	16	0	8	0	1	0	0	
<div>🛒</div>	000d566c-96c7-4f1c-...	TCGA-PRAD	Prostate	Male	428	2	390	5	5	16	0	8	1	1	0	0	
<div>🛒</div>	0011a67b-1ba9-4a32-...	TCGA-READ	Colorectal	Male	392	2	348	9	2	16	0	10	1	1	3	0	
<div>🛒</div>	001887aa-36d0-463f-...	TCGA-LIHC	Liver	Male	52	2	6	6	5	24	0	8	0	1	0	0	
<div>🛒</div>	001944e5-af34-4061-...	TCGA-BLCA	Bladder	Male	46	2	5	6	8	16	0	8	0	1	0	0	
<div>🛒</div>	001ad307-4ad3-4f1d-...	TCGA-LGG	Brain	Male	434	2	392	7	7	16	0	8	1	1	0	0	
<div>🛒</div>	001cef41-ff86-4d3f-a1...	TCGA-BRCA	Breast	Female	43	2	5	6	2	16	0	11	0	1	0	0	
<div>🛒</div>	001e0309-9c50-42b0-...	TCGA-UCEC	Uterus	Female	387	2	338	11	2	18	0	11	1	1	3	0	
<div>🛒</div>	0020317d-d10e-4e75-...	TCGA-SARC	Mesenchymal	Female	423	2	390	3	2	16	0	8	1	1	0	0	
<div>🛒</div>	0022478c-4dfd-4cbe-...	TCGA-KIRC	Kidney	Male	52	2	8	9	4	16	0	11	0	2	0	1	
<div>🛒</div>	0024ab57-4036-4b0f-...	TCGA-PCPG	Nervous System	Male	428	2	390	5	5	16	0	8	1	1	0	0	
<div>🛒</div>	002724fa-7051-49fa-9...	TCGA-UCEC	Uterus	Female	387	2	338	11	2	18	0	11	1	1	3	0	
<div>🛒</div>	0030a28c-81aa-44b0-...	TCGA-DLBC	Lymph Nodes	Female	425	2	390	5	2	16	0	8	1	1	0	0	
<div>🛒</div>	00397ce0-a931-46dd-...	TCGA-MESO	Mesothelium	Female	39	2	5	5	2	16	0	8	0	1	0	0	
<div>🛒</div>	0044844b-1125-5992-...	TARGET-WT	Kidney	Female	28	1	12	4	2	2	0	6	0	1	0	0	
<div>🛒</div>	0045349c-69d9-4306-...	TCGA-BRCA	Breast	Female	43	2	5	6	2	16	0	11	0	1	0	0	
<div>🛒</div>	0050d8be-1db6-4c17-...	TCGA-THYM	Thymus	Female	439	2	390	5	16	16	0	8	1	1	0	0	
<div>🛒</div>	005669e5-1a31-45fb-...	TCGA-LUSC	Lung	Male	39	2	5	5	2	16	0	8	0	1	0	0	
<div>🛒</div>	005a6517-2e5a-4ea3-...	TCGA-OV	Ovary	Female	422	2	357	10	8	26	2	15	1	1	0	0	
<div>🛒</div>	0061e9ef-df6a-414e-a...	TCGA-LGG	Brain	Male	430	2	386	8	7	16	0	9	1	1	0	0	

Show

20 ▴

entries

«

<

1

2

3

4

5

6

7

8

9

10

...

>





























































»

Clicking on any of the numbers in the table will apply this parameter to the filter currently being applied.

Cart	Case ID	Project	Primary Site	Gender	Files	Clinical	Array	Seq	Available
	0004d251-3f70-439...	TCGA-LIHC	Liver	Male	39	2	5	5	
Add all Case files to the Cart (39)			Prostate	Male	428	2	390	5	
	0011a67b-1ba9-4a3...	TCGA-READ	Colorectal	Male	392	2	348	9	
	001887aa-36d0-463...	TCGA-LIHC	Liver	Male	52	2	6	6	
	001944e5-af34-406...	TCGA-BLCA	Bladder	Male	46	2	5	6	

From this tab, clicking on the shopping cart will add all files related to this case to the cart.

Files

Summary Cases (11,484) Files (578,904)								
Files								
Showing 1 - 20 of 578,904 files								
	Access	File Name	Cases	Project	Data Type	Data Format	Size	Annotations
 	 Controlled	000aa811c15656604...	68c229a9-4c98-495a...	TCGA-SARC	Raw sequencing data	BAM	6.64 GB	0
 	 Controlled	0017ba4c33a07ba80...	70fc011c-0f22-4026-...	TCGA-BRCA	Raw sequencing data	BAM	12.08 GB	1
 	 Controlled	00286ada95aed3619...	7a077812-f6ee-4820-...	TCGA-UCEC	Raw sequencing data	BAM	8.92 GB	1
 	 Controlled	003860a34c9b244a5...	843c8f16-5e59-40d4-...	TCGA-BRCA	Raw sequencing data	BAM	7.71 GB	0
 	 Controlled	0047eb6e338c99ed7...	3e637872-f5e5-49d3-...	TCGA-BRCA	Raw sequencing data	BAM	4.38 GB	0
 	 Controlled	0048523803f0838a87...	b10c64c2-7fd2-4210-...	TCGA-CHOL	Raw sequencing data	BAM	5.31 GB	0
 	 Controlled	00586a9e7ba372bfa7...	76f98002-e8df-4dfa-...	TCGA-UCEC	Raw sequencing data	BAM	7.04 GB	0
 	 Controlled	006178ba37345d0e4...	3c0bb5e2-ad33-4d91...	TCGA-ESCA	Raw sequencing data	BAM	5.45 GB	0
 	 Controlled	00649547f0ace32d81...	7486a2b3-b09f-4e9c-...	TCGA-BRCA	Raw sequencing data	BAM	9.64 GB	0
 	 Controlled	00925769611d88cb0...	5362cbc4-cd28-4afd-...	TCGA-BRCA	Raw sequencing data	BAM	7.42 GB	0
 	 Controlled	009376b215547fc1ca...	3e240dc0-80b7-41ea...	TCGA-UCEC	Raw sequencing data	BAM	4.29 GB	0
 	 Controlled	00a016cbe7ea29145...	7f9031da-124a-4a38-...	TCGA-SARC	Raw sequencing data	BAM	82.02 GB	0
 	 Controlled	00a0680ba9b25a87df...	13f31bcb-befd-4ed2-...	TCGA-UCEC	Raw sequencing data	BAM	12.00 GB	0
 	 Controlled	00a0b7fe73741312c2...	5fd37868-4762-4109-...	TCGA-BRCA	Raw sequencing data	BAM	12.50 GB	0
 	 Controlled	00b34ca7322b3318d...	ca0f032b-a34c-4a92-...	TCGA-SARC	Raw sequencing data	BAM	6.72 GB	1
 	 Controlled	00c45a7d453026714...	9205c164-7975-421a...	TCGA-UCEC	Raw sequencing data	BAM	586 MB	0
 	 Controlled	00cb89ed6515afb666...	80809d79-92e1-4d6c...	TCGA-COAD	Raw sequencing data	BAM	10.03 GB	0
 	 Controlled	00cdc87c8d513c944f...	b1b9173c-b370-4fe5...	TCGA-UCEC	Raw sequencing data	BAM	4.39 GB	0
 	 Controlled	00cff657fd1d54444...	180a2b9c-9cac-4d96...	TCGA-CHOL	Raw sequencing data	BAM	8.00 GB	0
 	 Controlled	00f51fb4a56d82a3c9...	1a094c2b-9fba-4bb1...	TCGA-UCEC	Raw sequencing data	BAM	5.43 GB	0
Show 20 entries								
« ‹ 1 2 3 4 5 6 7 8 9 10 ... › »								

The Files tab provides a list of all cases available on the GDC Data Portal. As with other sections of the portal, results can be filtered down via facets. Looking at the query field on top of the page is the best way to easily identify if a filter is currently being applied.

Three actions are available from this tab.

Summary

Cases (11,484)

Files (578,904)

Files

Showing 1 - 20 of 578,904 files

	Access	File Name	Case
<div> <div></div> <div>Add all files to the Cart</div> </div> <div> <div></div> <div>Remove all from the Cart</div> </div>		5656604...	68c2
		3a07ba80...	70fc
		5aed3619...	7a07
<div> <div></div> <div></div> </div> <div> <div></div> <div></div> </div> <div> <div></div> <div>Controlled</div> </div>		003860a34c9b244a5...	843c
<div> <div></div> <div></div> </div> <div> <div></div> <div></div> </div> <div> <div></div> <div>Controlled</div> </div>		0047eb6e338c99ed7...	3e63
<div> <div></div> <div></div> </div> <div> <div></div> <div></div> </div> <div> <div></div> <div>Controlled</div> </div>		0048523803f0838a87...	b10c

In the table header clicking on the shopping cart will give user the option to have all files matching current filter to be either added or removed from the cart.

Alternatively, for each row user can:

- Add the file to the cart
- Download the file directly

Advanced

The ‘*Advanced*’ button redirects to the advanced query page; this feature is detailed in the Advanced Search section of the documentation.

Case Entity Page

 7f9031da-124a-4a38-83e6-878a50e58c24

Summary

Case UUID	7f9031da-124a-4a38-83e6-878a50e58c24
Case Barcode	TCGA-FX-A3RE
Project ID	TCGA-SARC
Project Name	Sarcoma
Disease Type	Sarcoma
Program	TCGA
Primary Site	Mesenchymal
Status	Alive
Number of Data Files	440
Number of Annotations	0

Experimental Strategy

Analysis	Files
Wxs	2
Genotyping Array	20
Dna-Seq	12
Mirna-Seq	4
Protein Expression Array	386
Methylation Array	4
Rna-Seq	8
Wgs	2

Available Data

Data Type	Files
Clinical	2
Raw Microarray Data	390
Raw Sequencing Data	8
Simple Nucleotide Variation	14
Copy Number Variation	16
Structural Rearrangement	0
Gene Expression	8
Protein Expression	1
Dna Methylation	1
Other	0

The Case Entity page displays case details including the project and disease information, data files that are available for that case, and the experimental strategies employed.

Clinical

Download Clinical XML

Gender	Female
Age at Diagnosis	
Vital Status	Alive
Days to death	
ICD10	C56.9
Ethnicity	Not Hispanic Or Latino
Race	White

Biospecimen

Download Biospecimen XML

Samples

TCGA-FX-A3RE-01A

Portions

TCGA-FX-A3RE-01A-11

Slides

TCGA-FX-A3RE-01A-01-TS1

Analytes

TCGA-FX-A3RE-01A-21-A455-20

TCGA-FX-A3RE-10A

Details - Slide

Submitter ID	TCGA-FX-A3RE-01A-01-TS1
Slide ID	abcb4233-9d30-4fe8-a0f8-f2e2caaa77b2
Percent Tumor Nuclei	100
Percent Monocyte Infiltration	0
Percent Normal Cells	0
Percent Stromal Cells	5
Percent Eosinophil Infiltration	--
Percent Lymphocyte Infiltration	0
Percent Neutrophil Infiltration	0
Section Location	TOP
Percent Granulocyte Infiltration	--
Percent Necrosis	0
Percent Inflammation Infiltration	--
Number Proliferating Cells	--
Percent Tumor Cells	95

The Case Entity page also provides direct access to the Clinical information about that case, and the Biospecimen file associated with the samples collected from that case (if present). These files can be downloaded using the links in this view.

File Entity Page

Clicking on a filename on the GDC Data Portal redirects to a dedicated page providing additional details about the file. From there the file can be added to cart, downloaded directly, or if the file is a BAM, just a slice can be downloaded.

35

fd2386b3-32f5-4aea-a55c-af24f5d1e650

[Add to Cart](#) [Download](#) [BAM Slicing](#)

File Properties

File Name	000aa811c15656604161e8f0e3a0aae4.bam
File Access	Controlled
UUID	fd2386b3-32f5-4aea-a55c-af24f5d1e650
File Origin	Migrated
Data format	BAM
File Size	6.64 GB
Checksum	000aa811c15656604161e8f0e3a0aae4
Published	11/17/2015
Uploaded	11/17/2015
State	Live
Archive	--

Data Information

Data Type	Raw Sequencing Data
Data Subtype	Aligned Reads
Experimental Strategy	WXS
Platform	Illumina HiSeq
Data Submitter	WUSM
Tag	--

Associated Cases / Biospecimen

Entity ID	Entity Type	Case UUID	Annotations
52b82f28-4796-456a-8523-59c171707336	Aliquot	68c229a9-4c98-495a-b75a-f6f2172b075b	0

Related Files

File Name	File Type	Checksum	Action
000aa811c15656604161e8f0e3a0aae4.ba...	bai	54bcf9e1be6c662a31aeb88e59cf99e1	Download

Relation between this file and its case as well as between this file and any other file is also available in this page.

If the file was originally produced as part of an archive containing many other files, the archive ID is displayed and, if selected, the user will see a 'Files' view containing all other files in that archive.

If the file was part of a MAGE-TAB archive, the whole archive can be downloaded from the file details page. BAI sequence data index files are related to sequence data BAM files and will be displayed if relevant.

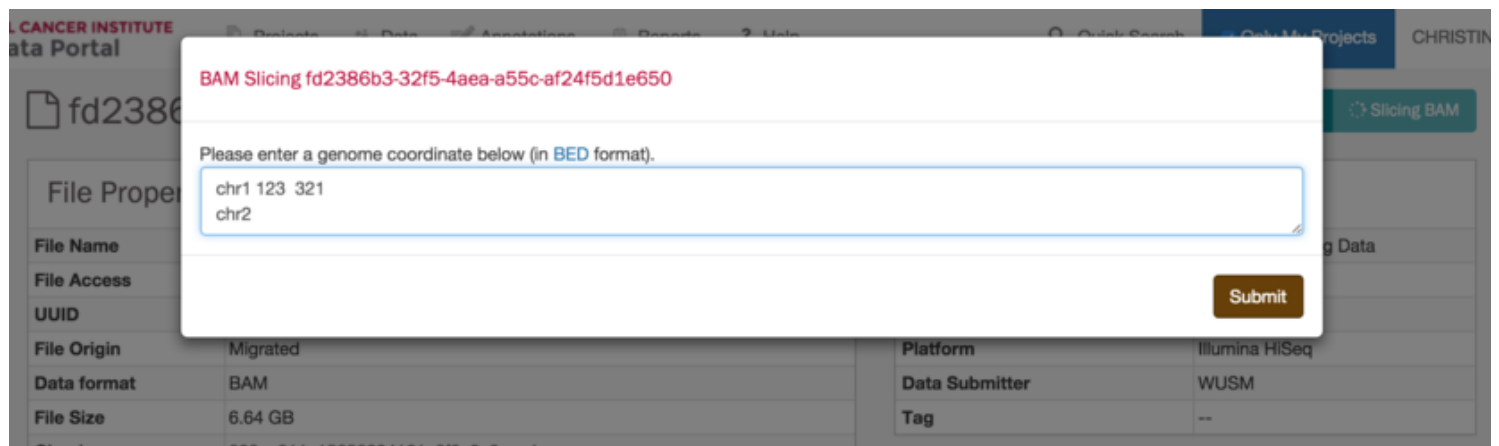
Chapter 7

BAM Slicing

Overview

! Section to be further completed !

If the file listed in the entity page is a BAM file, a “BAM Slicing” button will be available. Clicking on it will open the BAM slicing window.



The screenshot shows a web application interface for the Cancer Institute Data Portal. A modal window titled "BAM Slicing fd2386b3-32f5-4aea-a55c-af24f5d1e650" is open. It contains a text input field with the placeholder "Please enter a genome coordinate below (in BED format)." and a "Submit" button. The background shows a file details page for a BAM file with the following properties:

File Properties	
File Name	fd2386b3-32f5-4aea-a55c-af24f5d1e650
File Access	Public
UUID	fd2386b3-32f5-4aea-a55c-af24f5d1e650
File Origin	Migrated
Data format	BAM
File Size	6.64 GB
Platform	Illumina HiSeq
Data Submitter	WUSM
Tag	--

During preparation of the slice, the icon on the BAM Slicing button will be spinning, and the file will be offered for download to the user as soon as ready.

Chapter 8

Annotations

Annotations Search

Overview

Annotations are metadata, notes and caveats referring to cases, samples or files after curation.

Annotations are important as they indicate that issues or caveats have been found with the cases, samples, or data after it has been submitted.

Annotation

Search for Annotation ID

Primary Site

Ovary

1,662

Colorectal

1,762

Brain

1,688

Lung

1,513

Stomach

21 More...

Project

TCGA-OV

10,701

TCGA-STAD

1,513

TCGA-COAD

1,466

TCGA-BRCA

1,452

TCGA-GBM

1,344

27 More...

Entity Type

Aliquot

21,211

Participant

1,343

Portion

967

Analyte

672

Sample

92

Annotation Category

Item Flagged DNU

20,900

Item Is Noncanonical

1,526

Prior Malignancy

403

General

307

Start searching by selecting a facet

Annotations

Showing 1 - 20 of 24,285 annotations

ID	Case ID	Project	Entity Type	Entity ID	Category	Classification	Created Date	Annotator
01deb78b-0560-5b45...	6209e80d-a115-4f8e...	TCGA-OV	aliquot	88656b94-cd24-4c4c...	Item flagged DNU	CenterNotification	07/30/2012	DCC
b52d4a4c-0669-5b3f...	92badeb5-a50e-4a62...	TCGA-OV	aliquot	d525e6af-f667-4891-...	Item flagged DNU	CenterNotification	01/28/2015	DCC
952e125a-fd44-58c8...	62d32d19-ec93-4359...	TCGA-STAD	aliquot	9c2721cf-ca2e-4a60-...	Item flagged DNU	CenterNotification	08/13/2014	DCC
5d1017d9-fcfa-59b8...	dac96910-0cb2-44ea...	TCGA-OV	aliquot	b75db551-4254-4736...	Item flagged DNU	CenterNotification	01/28/2015	DCC
261b4621-127f-5f37...	8f3acbbd-b2ca-4e66...	TCGA-GBM	aliquot	6fad6fd1-5709-4451-...	Item flagged DNU	CenterNotification	09/28/2012	DCC
7e456d04-92da-529d...	6d10d4ee-6331-4bba...	TCGA-OV	aliquot	7fbfdb3e-1fd2-4206...	Item flagged DNU	CenterNotification	02/19/2015	DCC
716186c5-3b94-5553...	dfa3ef71-7006-4c6d...	TCGA-GBM	aliquot	9d45de65-47f2-4482...	Item flagged DNU	CenterNotification	10/02/2013	DCC
ad6ebe37-d99b-5b11...	9c48646e-7f66-4936...	TCGA-BLCA	aliquot	ad54389a-c08c-451f...	Item flagged DNU	CenterNotification	02/04/2014	DCC
5dc0b127-78e7-5038...	0de910b1-edaf-47e4...	TCGA-OV	aliquot	9bd35257-52b8-4277...	Item flagged DNU	CenterNotification	01/28/2015	DCC
834afde2-a0c0-5d9b...	5d603bba-62e6-48fb...	TCGA-OV	aliquot	fc32468d-de77-4e0e...	Item flagged DNU	CenterNotification	02/19/2015	DCC
525f01a7-7ce2-576c...	62d32d19-ec93-4359...	TCGA-STAD	aliquot	9c2721cf-ca2e-4a60...	Item flagged DNU	CenterNotification	08/13/2014	DCC
33f38052-bf22-59f1...	e2f6d3aa-b817-4193...	TCGA-LIHC	aliquot	ac97da22-e85c-41e3...	Item flagged DNU	CenterNotification	09/11/2011	DCC
785c3503-083a-516b...	b46263ab-c3ca-4fda...	TCGA-OV	aliquot	08e7c700-7038-40b5...	Item flagged DNU	CenterNotification	11/30/2012	DCC
0c097b40-4469-59d7...	40f43483-b504-4ea7...	TCGA-UCEC	aliquot	7e8e52bd-778b-4f44...	Item flagged DNU	CenterNotification	04/28/2011	DCC
b85c55cf-658e-59c5...	b1e11e94-646b-4c44...	TCGA-OV	aliquot	5345eef4-ee81-4128...	Item flagged DNU	CenterNotification	07/09/2012	DCC
49c61840-df0f-5a22...	b25b3379-df8f-4900...	TCGA-OV	aliquot	0dc4136c-c8d3-43b3...	Item flagged DNU	CenterNotification	08/21/2012	DCC
d8b87bd3-9e9e-59a4...	97ad0360-3be4-401a...	TCGA-OV	aliquot	e39a23c8-1bc3-4bfb...	Item flagged DNU	CenterNotification	12/24/2012	DCC
538ed497-0238-5a27...	680f23a9-fcce-493e...	TCGA-BRCA	aliquot	8ff00a87-b2bd-4ff3-9...	Item flagged DNU	CenterNotification	10/10/2013	DCC
9cafa76d-3389-5f03...	9a30dee6-962a-49db...	TCGA-OV	aliquot	639429a9-c1d3-410d...	Item flagged DNU	CenterNotification	02/19/2015	DCC
49fda51c-d558-5676...	0ea4920f-f4c8-4590...	TCGA-LUAD	aliquot	2961c12e-ca46-42a3...	Item flagged DNU	CenterNotification	01/09/2013	DCC

Show

20

entries

«

<

1

2

3

4

5

6

7

8

9

10

...

>

»

Facets

The following facets are available to search for annotations:

- **Annotation ID:** Seach using annotation ID
- **Primary Site:** Tissue type of the cancer

- **Project:** See project section of the documentation
- **Entity Type:** Patient, Sample, Portion, Slide, Analyte, Aliquot
- **Annotation Category:** See link below
- **Annotation Created Since:** Search for annotations by date of creation
- **Annotation Classification:** See link below

Annotation List

The list of annotations is displayed on the right side of the page, clicking on an Annotation ID will redirect the the annotation entity page

Annotation Categories and Classification

For more details about categories and classifications please refer to the [TCGA Annotations page on NCI Wiki](#)

Annotation Entity Page

The annotation entity page provides more details about a specific annotation. It is available by clicking on an Annotation ID somewhere on the portal.



01deb78b-0560-5b45-94af-e34611f1c666

Summary

Annotation ID	01deb78b-0560-5b45-94af-e34611f1c666
Entity ID	88656b94-cd24-4c4c-b625-635ef692cf13
Entity Barcode	TCGA-23-1113-01A-01R-1564-13
Entity Type	Aliquot
Case ID	6209e80d-a115-4f8e-bf0d-18461401a1c6
Project ID	TCGA-OV
Classification	Center Notification
Category	Item flagged DNU
Created On	07/30/2012
Created By	DCC
Status	Approved

Notes

SDRF in bcgsc.ca_OV.IlluminaHiSeq_RNASeq.mage-tab.1.8.0 flagged aliquot to be excluded for analysis based on file 'TCGA-23-1113-01A-01R-1564-13.exon.quantification.txt'.

Chapter 9

Advanced Search

Overview

An advanced search allows, via Genomic Query Language (GQL), to use structured queries to search for files and cases.

Valid Query

Start Typing Your Query....

Basic Reset Q Search


Summary

Cases (11,484)


Files (581,003)

🛒 Add all files to the Cart


FILES
581,003



CASES
11,484



FILE SIZE
1.22 PB



The advanced search provides more filters than the faceted search.

A simple query in GQL (also known as a ‘clause’) consists of a **field**, followed by an **operator**, followed by one or more **values**. For example, the following simple query will find all files in the “TEST” project:

```
cases.project.primary_site = Brain
```

Valid Query

participants.project.primary_site = Brain

Basic Reset

Q Search

Summary

Cases (1,133)

Files (57,646)

Add all files to the Cart

FILES

57,646

CASES

1,133

FILE SIZE

114.99 TB

Be aware that it is not possible to compare two fields (e.g. disease_type = project.name).

Note: GQL is not a database query language. For example, GQL does not have a **SELECT** statement.

Switch between ‘Advanced’ and ‘Faceted’ Search

A query created using ‘Faceted Search’ will be automatically translated to ‘Advanced Search’ (i.e. GQL) when clicking on the “Advanced” button.

However, a query created using ‘Advanced Search’ is not translated to ‘Faceted Search’. From the advanced search, click on Basic and you will be redirected to a reset faceted search.

Use the Advanced Search

When opening the advanced search page (via the Data view), the search field will be automatically populated with facets filters already applied (if any).

This default query can be removed by pressing “Reset”.

Once the query has been entered and is identified as a “Valid Query”, click on “Search” to run your query.

Auto-complete

As a query is being written, the GDC Data Portal will analyse the context and offer a list of “auto-complete”.

Field Auto-complete

The list of auto-complete suggestions includes **all** the fields authorized in the advanced search matching the user text input. The user has to scroll down to see more fields in the dropdown.

Invalid Query (see errors)

project|

project.disease_type

project.name

project.primary_site

project.program.name

project.program.program_id

project.project_id

project.state

tissue_source_site.project

Basic

Reset

Q Search

Files (581,003)

CASES

11,484

FILE SIZE

1.22 PB

Value Auto-complete

The list of auto-complete suggestions includes top 100 values that match the user text input. The user has to scroll down to see more values in the dropdown.

The value auto-complete is not aware of the general context of the query, the system will display all available values in GDC for the selected field. It means the query could return 0 result depending of other filters.

Invalid Query (see errors)

participants.project.name = |

☐

Breast Invasive Carcinoma

☐

Glioblastoma Multiforme

☐

Ovarian Serous Cystadenocarcinoma

☐

Lung Adenocarcinoma

☐

Uterine Corpus Endometrial Carcinoma

☐

Kidney Renal Clear Cell Carcinoma

☐

Head and Neck Squamous Cell Carcinoma

☐

Brain Lower Grade Glioma

☐

Lung Squamous Cell Carcinoma


☐

Thyroid Carcinoma

BasicReset


Search

34



FILE SIZE

1.22 PB



File Counts by Primary Site

Note: Quotes are automatically added to the value if it contains spaces.

Setting Precedence of Operators

You can use parentheses in complex GQL statements to enforce the precedence of operators.

For example, if you want to find all the open files in TCGA program as well as the files in TARGET program, you can use parentheses to enforce the precedence of the boolean operators in your query, i.e.:

```
(files.access = open and cases.project.program.name = TCGA) or cases.project.program.name = TARGET
```

Note: Without parentheses, the statement will be evaluated left-to-right.

Keywords

A GQL keyword is a word that joins two or more clauses together to form a complex GQL query.

List of Keywords:

- AND
- OR

Note: parentheses can be used to control the order in which clauses are executed.

AND Keyword

Used to combine multiple clauses, allowing you to refine your search.

Examples:

- Find all open files in breast cancer:

```
cases.project.primary_site = Breast and files.access = open
```

- Find all open files in breast cancer and data type is copy number variation:

```
cases.project.primary_site = Breast and files.access = open and files.data_type = "Copy number variation"
```

OR Keyword

Used to combine multiple clauses, allowing you to expand your search.

Note: IN keyword can be an alternative to OR and result in simplified queries.

Examples:

- Find all files that are raw sequencing data or raw microarray data:

```
files.data_type = "Raw microarray data" or files.data_type = "Raw sequencing data"
```

- Find all files where donors are male or vital status is alive:

```
cases.clinical.gender = male or cases.clinical.vital_status = alive
```

Operators

An operator in GQL is one or more symbols or words comparing the value of a field on its left with one or more values on its right, such that only true results are retrieved by the clause.

List of Operators:

Operator	Name
=	EQUAL
!=	NOT EQUAL
<	LOWER THAN
<=	LOWER THAN OR EQUAL
>	GREATER THAN
>=	GREATER THAN OR EQUAL
IN	IN
NOT IN	NOT IN
IS MISSING	IS MISSING
IS NOT MISSING	IS NOT MISSING

Limitation: <, <=, >, >=, Cannot be used with **text** or **date** fields.

“=” operator - EQUAL

The “=” operator is used to search for files where the value of the specified field exactly matches the specified value.

Examples:

- Find all files that are gene expression:

```
files.data_type = "Gene expression"
```

- Find all cases whose gender is female:

```
cases.clinical.gender = female
```

“!=” operator - NOT EQUAL

The “!=” operator is used to search for files where the value of the specified field does not match the specified value.

The “!=” operator will not match a field that has no value (i.e. a field that is empty). For example, `gender != male` will only match cases who have a gender and the gender is not male. To find cases other than male or with no gender populated, you would need to type `gender != male` or `gender` is missing.

Examples:

- Find all files with an experimental different from genotyping array:

```
files.experimental_strategy != "Genotyping array"
```

“>” operator - GREATER THAN

The “>” operator is used to search for files where the value of the specified field is greater than the specified value.

Examples:

- Find all cases whose number of days to death is greater than 60:

```
cases.clinical.days_to_death > 60
```

- Find all cases whose year of diagnosis is after 2000:

```
cases.clinical.year_of_diagnosis > 2000
```

“>=” operator - GREATER THAN OR EQUALS

The “>=” operator is used to search for files where the value of the specified field is greater than or equal to the specified value.

Examples:

- Find all cases whose number of days to death is equal or greater than 60:

```
cases.clinical.days_to_death >= 60
```

- Find all cases whose year of diagnosis is in 2005 or later:

```
cases.clinical.year_of_diagnosis >= 2005
```

“<” operator - LESS THAN

The “<” operator is used to search for files where the value of the specified field is less than the specified value.

Examples:

- Find all cases whose age at diagnosis is less than 400 days:

```
cases.clinical.age_at_diagnosis < 400
```

- Find all cases whose year of diagnosis is before 1990:

```
cases.clinical.year_of_diagnosis < 1990
```

“<=” operator - LESS THAN OR EQUALS

The “<=” operator is used to search for files where the value of the specified field is less than or equal to the specified value.

Examples:

- Find all cases whose year of diagnosis is in 1990 or earlier:

```
cases.clinical.year_of_diagnosis <= 1990
```

- Find all cases with a number of days to death less than or equal to 20:

```
cases.clinical.days_to_death <= 20
```

“in” Operator

The “IN” operator is used to search for files where the value of the specified field is one of multiple specified values. The values are specified as a comma-delimited list, surrounded by brackets [].

Using “IN” is equivalent to using multiple **EQUALS** (=) statements, but is shorter and more convenient. That is, typing **project IN [ProjectA, ProjectB, ProjectC]** is the same as typing **project = "ProjectA" OR project = "ProjectB" OR project = "ProjectC"**.

Examples:

- Find all files in breast, breast and lung and cancer:

```
cases.project.primary_site in [Brain, Breast,Lung]
```

- Find all files tagged with exon or junction or hg19:

```
files.tags in [exon, junction, hg19]
```

“not in” Operator

The " NOT IN " operator is used to search for files where the value of the specified field is not one of multiple specified values.

Using “NOT IN” is equivalent to using multiple NOT_EQUALS (!=) statements, but is shorter and more convenient. That is, typing `project NOT IN [ProjectA, ProjectB, ProjectC]` is the same as typing `project != "ProjectA" OR project != "ProjectB" OR project != "ProjectC"`.

The “NOT IN” operator will not match a field that has no value (i.e. a field that is empty). For example, `experimental_strategy not in ["WGS","WXS"]` will only match files that have an experimental strategy **and** the experimental strategy is not “WGS” or “WXS”. To find files with an experimental strategy different from than “WGS” or “WXS” **or is not assigned**, you would need to type: `files.experimental_strategy in ["WXS","WGS"]` or `files.experimental_strategy is missing`.

Examples:

- Find all issues where experimental strategy is not WXS, WGS, Genotyping array:

```
files.experimental_strategy not in ["WXS","WGS", "Genotyping array"]
```

“is missing” Operator

The “IS” operator can only be used with missing. That is, it is used to search for files where the specified field has no value.

Examples:

- Find all cases where gender is missing:

```
cases.clinical.gender is missing
```

“is not missing” Operator

The " IS NOT " operator can only be used with missing. That is, it is used to search for files where the specified field has a value.

Examples:

- Find all cases where race is not missing:

```
cases.clinical.race is not missing
```

Special Cases

Using Quotes

A value must be quoted if it contains a space. Otherwise the advanced search will not be able to interpret the value.

Quotes are not necessary if the value consists of one single word.

- Example: Find all cases with primary site is brain and data type is copy number variation:

```
cases.project.primary_site = Brain and files.data_type = "Copy number variation"
```


Age at Diagnosis - Unit in Days

The unit for age at diagnosis is in **days**. The user has to convert the number of years to number of days.

- Example: Find all cases whose age at diagnosis > 40 years old (40 * 365)

```
cases.clinical.age_at_diagnosis > 14600
```

Fields not Supported in the Advanced Search

The dates fields are currently not supported in the advanced search.

For example, a user cannot look for all the files which were created after January 1, 2015 (creation_datetime).

Fields Reference

The full list of fields available on the GDC Data Portal can be found through the GDC API using the following endpoint:

https://gdc-api.nci.nih.gov/gql/_mapping

Alternatively, a static list of those fields is available below.

Files

- files.access
- files.acl
- files.archive.archive_id
- files.archive.revision
- files.archive.submitter_id
- files.center.center_id
- files.center.center_type
- files.center.code
- files.center.name
- files.center.namespace
- files.center.short_name
- files.data_format
- files.data_subtype
- files.data_type
- files.experimental_strategy
- files.file_id
- files.file_name
- files.file_size
- files.md5sum
- files.origin
- files.platform
- files.related_files.file_id
- files.related_files.file_name
- files.related_files.md5sum
- files.related_files.type
- files.state
- files.state_comment
- files.submitter_id
- files.tags

Cases

- cases.clinical.age_at_diagnosis
- cases.clinical.clinical_id
- cases.clinical.days_to_death
- cases.clinical.ethnicity
- cases.clinical.gender
- cases.clinical.icd_10
- cases.clinical.race
- cases.clinical.vital_status
- cases.clinical.year_of_diagnosis
- cases.case_id
- cases.project.disease_type
- cases.project.name
- cases.project.program.name
- cases.project.program.program_id
- cases.project.project_id
- cases.project.state
- cases.samples.current_weight
- cases.samples.days_to_collection
- cases.samples.days_to_sample_procurement
- cases.samples.freezing_method
- cases.samples.initial_weight
- cases.samples.intermediate_dimension
- cases.samples.is_ffpe
- cases.samples.longest_dimension
- cases.samples.oct_embedded
- cases.samples.pathology_report_uuid
- cases.samples.portions.analytes.a260_a280_ratio
- cases.samples.portions.analytes.aliquots.aliquot_id
- cases.samples.portions.analytes.aliquots.amount
- cases.samples.portions.analytes.aliquots.center.center_id
- cases.samples.portions.analytes.aliquots.center.center_type
- cases.samples.portions.analytes.aliquots.center.code
- cases.samples.portions.analytes.aliquots.center.name
- cases.samples.portions.analytes.aliquots.center.namespace
- cases.samples.portions.analytes.aliquots.center.short_name
- cases.samples.portions.analytes.aliquots.concentration
- cases.samples.portions.analytes.aliquots.source_center
- cases.samples.portions.analytes.aliquots.submitter_id
- cases.samples.portions.analytes.amount
- cases.samples.portions.analytes.analyte_id
- cases.samples.portions.analytes.analyte_type
- cases.samples.portions.analytes.concentration
- cases.samples.portions.analytes.spectrophotometer_method
- cases.samples.portions.analytes.submitter_id
- cases.samples.portions.analytes.well_number
- cases.samples.portions.center.center_id
- cases.samples.portions.center.center_type
- cases.samples.portions.center.code
- cases.samples.portions.center.name
- cases.samples.portions.center.namespace
- cases.samples.portions.center.short_name


- cases.samples.portions.is_ffpe
- cases.samples.portions.portion_id
- cases.samples.portions.portion_number
- cases.samples.portions.slides.number_proliferating_cells
- cases.samples.portions.slides.percent_eosinophil_infiltration
- cases.samples.portions.slides.percent_granulocyte_infiltration
- cases.samples.portions.slides.percent_inflam_infiltration
- cases.samples.portions.slides.percent_lymphocyte_infiltration
- cases.samples.portions.slides.percent_monocyte_infiltration
- cases.samples.portions.slides.percent_necrosis
- cases.samples.portions.slides.percent_neutrophil_infiltration
- cases.samples.portions.slides.percent_normal_cells
- cases.samples.portions.slides.percent_stromal_cells
- cases.samples.portions.slides.percent_tumor_cells
- cases.samples.portions.slides.percent_tumor_nuclei
- cases.samples.portions.slides.section_location
- cases.samples.portions.slides.slide_id
- cases.samples.portions.slides.submitter_id
- cases.samples.portions.submitter_id
- cases.samples.portions.weight
- cases.samples.sample_id
- cases.samples.sample_type
- cases.samples.sample_type_id
- cases.samples.shortest_dimension
- cases.samples.submitter_id
- cases.samples.time_between_clamping_and_freezing
- cases.samples.time_between_excision_and_freezing
- cases.samples.tumor_code
- cases.samples.tumor_code_id
- cases.submitter_id
- cases.tissue_source_site.bcr_id
- cases.tissue_source_site.code
- cases.tissue_source_site.name
- cases.tissue_source_site.project
- cases.tissue_source_site.tissue_source_site_id

Chapter 10

Reports

Data Download Statistics Report

The “Reports” section hosts various reports summarizing GDC Data Portal use and download statistics for various fields and Facets.

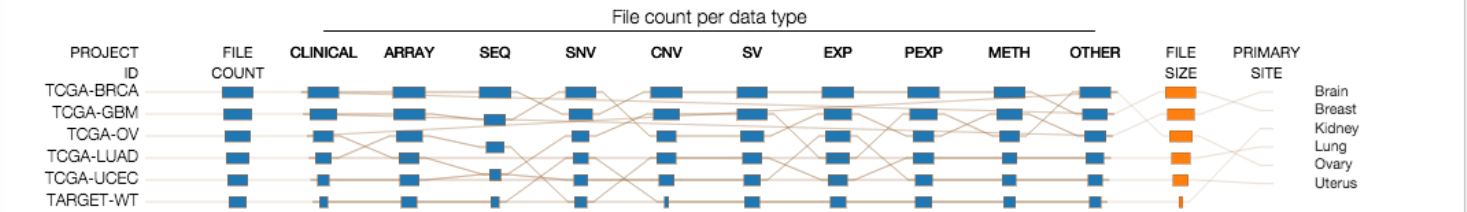
 **Data download statistics**

This report depicts statistics describing data download since the GDC Data Portal launch. Represented GDC data includes information about the total number and size of downloaded files categorized at project, data and user access levels.

Export Report

Project

Downloads by Project and Data Type



Downloads by Project

Project ID	Project Name	# of files	File Size
TCGA-BRCA		10,145,332	728.75 TB
TCGA-GBM		9,145,332	628.75 TB
TCGA-OV		8,145,332	528.75 TB
TCGA-LUAD		7,145,332	428.75 TB
TCGA-UCEC		6,145,332	328.75 TB
TARGET-WT		5,145,332	28.75 TB

The information contained in the reports can be downloaded in JSON format from the reports page.

Chapter 11

Cart and Files Download

Overview

While browsing through the GDC Data Portal, files can be downloaded either individually or bundled through a cart mechanism similar to online stores.

The cart pages shows a summary of all files currently in the cart, display a file count by project and identified the authorization level of those files (controlled or open access).

GDC Cart

ADD MORE DETAILS

Download Options

From the cart, three different download options are available to the end user.

Note: when downloading multiple files from the GDC Data Portal, those files are automatically bundled-up into one single Gzipped (.tar.gz) file.

GDC Data Transfer Tool

The “Manifest for Fast Download Tool” button will download a manifest that can be imported into the GDC Data Transfer Tool.

id	filename	md5	size	state
4ea9c657-8f85-44d0-9a77-ad59cced8973	mdanderson.org_ESCA.MDA_RPPA_Core.mage-tab.1.1.0.tar.gz \N	2516051	live	
b8342cd5-330e-440b-b53a-1112341d87db	mdanderson.org_SARC.MDA_RPPA_Core.mage-tab.1.1.0.tar.gz \N	4523632	live	
c57673ac-998a-4a50-a12b-4cac5dc3b72e	mdanderson.org_KIRP.MDA_RPPA_Core.mage-tab.1.2.0.tar.gz \N	4195746	live	
3f22dd8d-59c8-43a4-89cf-3b595f2e5a06	14-3-3_beta-R-V_GBL1112940.tif	56df0e4b4fc092fc3643bd2e316ac05b	6257	
7ce05059-9197-4d38-830f-04356f5f851a	14-3-3_beta-R-V_GBL11066140.tif	6abfee483974bc2e61a37b5499ae9a07	6261	
8e00d22a-ca6f-4da8-a1c3-f23144cb21b7	14-3-3_beta-R-V_GBL1112940.tif	56df0e4b4fc092fc3643bd2e316ac05b	6257	
96487cd7-8fa8-4bee-9863-17004a70b2e9	14-3-3_beta-R-V_GBL1112940.tif	56df0e4b4fc092fc3643bd2e316ac05b	6257	

The Manifest contains a list of the file UUIDs in the cart and can be used together with the GDC Data Transfer Tool to download all files.

Information on the GDC Data Transfer Tool is available in the [GDC Data Transfer Tool User's Guide](#).

Cart Download

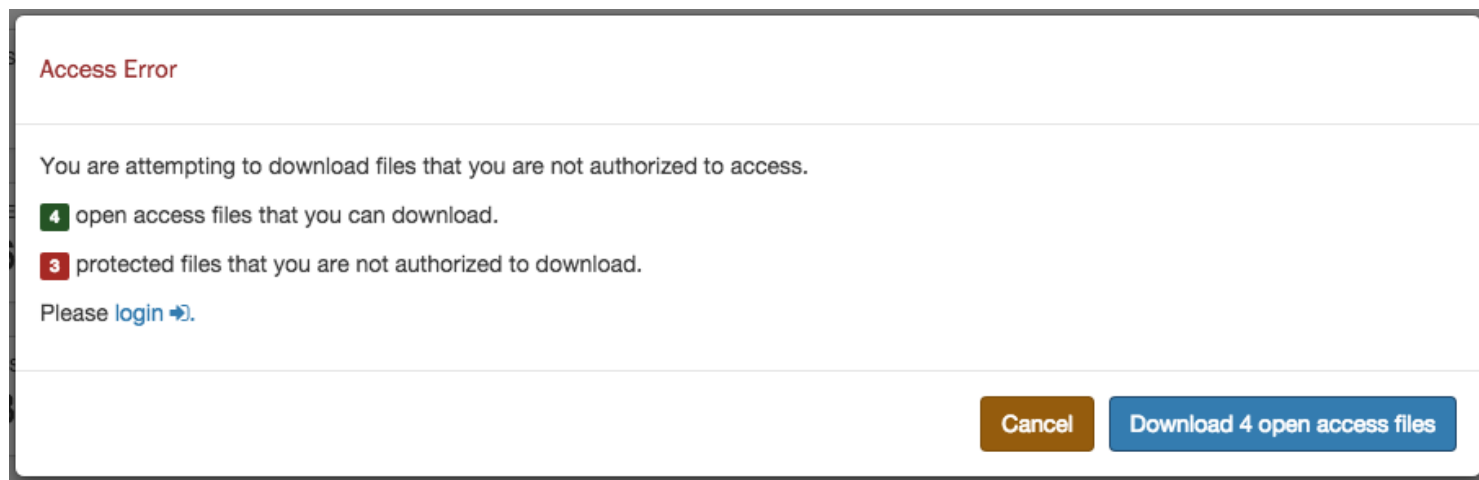
The “Download Cart (web browser)” button will initial a data download directly from the browser. Users have to be cautious of the amount of data in the cart since this option will not optimize bandwidth neither provide resume capabilities.

Individual Files Download

Similar to the files page, each row contains a download button to download a particular file individually.

Access

If a user tries to download a cart containing controlled data and without being authenticated, a pop-up will be displayed to offer the user either to download only open access files or to login into the GDC Data Portal through eRA Commons.



Chapter 12

Key Terms

The following table provides definitions and explanations for terms and acronyms relevant to the content presented within this document.

Term	Definition
API	Application Programming Interface
BAM	Binary Alignment/Map
BAI	Binary Sequence Alignment/Map Index
CCG	Center for Cancer Genomics
CGCI	Cancer Genome Characterization Initiative
CSV	Common Separated Values
dbGaP	The database of Genotypes and Phenotypes
eRA	Electronic Research Administration
GDC	Genomic Data Commons
HHS	Health & Human Services
HTTP	Hypertext Transfer Protocol
ID	Identifier
JSON	JavaScript Object Notation
MAGE-TAB	Microarray Gene Expression - Tabular format
miRNA	MicroRNA
NCI	National Cancer Institute
NIH	National Institutes of Health
TARGET	Therapeutically Applicable Research to Generate Effective Treatments
TCGA	The Cancer Genome Atlas
TSV	Tab Separated Values
UUID	Universally Unique Identifier
URL	Universal Resource Locator
VM	Virtual Machine
XML	eXtensible Markup Language

0.2.18.3

GDC Data Portal Release 0.2.18.3

Release Date: November 23, 2015

New Features and Changes

- BAM slicing UI in the File Entity page (only available for BAM files with BAI files)

Bugs Fixed Since Last Release

- Disease Type does not auto filter (Data Facet)

Known Issues and Workarounds

- Missing pathology files and slide images for some TCGA datasets since they are not connected to the biospecimen chain yet
- In some specific situation (complex queries using must_not term), the advanced search might returns incorrect results
- When exporting the annotation table to a file, the create date field is not displayed in the same format than the UI
- In the advanced search, adding a trailing space to the query causes the UI to display a syntax error although the query is valid
- Checksum missing for MAGE-TAB files
- User can download large files (> 1GB) from the Portal but it fails when extracting the tar.gz file. User should use the GDC Download Transfer Tool for large downloads
- Authenticated user with access to TARGET projects cannot download Target DCC files. They will get a pop-up message.
- In Case Entity Page, the “Download Clinical XML” button is not applicable to TARGET projects. The label is incorrect but the functionality is accurate.
- The advanced search suggests more fields than expected (e.g. “cases.aliqouts” should not be displayed). User should remove the field from the query and use the long field name (e.g. “cases.samples.portions.analytes.aliqouts”)
- If a user has previously logged into the Portal and left a session without logging out, if he comes back to the Portal after his sessionID expires, it looks like the user is still authenticated. The user cannot download the token and gets an error message that would not close. The user should clear the cache to properly log out.
- Sorting on size and removing files from cart does not work in IE 10
- Annotations page not section 508 compliant
- Facets are displayed above the results table when window is small
- Table sorting icon does not include numbers

Release details are maintained in the GDC Data Portal Change Log.

0.2.15.4

GDC Data Portal Release 0.2.15-oicr4

Release Date: October 1st, 2015

New Features and Changes

- n/a

Bugs Fixed Since Last Release

- Data access and subtype units are correct in the data download statistics report
- Improved 508 compliance of the portal (including the data download statistics report)
- Addressed an issue with the user getting an empty page in specific browsing situations
- “My projects” filter has been re-enabled for users authenticated with eRACommons but without dbGap authorization.

Known Issues and Workarounds

- Missing pathology files and slide images for some TCGA datasets since they are not connected to the biospecimen chain yet
- In some specific situation (complex queries using must_not term), the advanced search might returns incorrect results

- When exporting the annotation table to a file, the create date field is not displayed in the same format than the UI
- In the advanced search, adding a trailing space to the query causes the UI to display a syntax error although the query is valid
- Checksum missing for MAGE-TAB files
- User can download large files (> 1GB) from the Portal but it fails when extracting the tar.gz file. User should use the GDC Download Transfer Tool for large downloads
- Authenticated user with access to TARGET projects cannot download Target DCC files. They will get a pop-up message.
- In Case Entity Page, the “Download Clinical XML” button is not applicable to TARGET projects. The label is incorrect but the functionality is accurate.
- The advanced search suggests more fields than expected (e.g. “cases.aliqouts” should not be displayed). User should remove the field from the query and use the long field name (e.g. “cases.samples.portions.analytes.aliqouts”)
- If a user has previously logged into the Portal and left a session without logging out, if he comes back to the Portal after his sessionID expires, it looks like the user is still authenticated. The user cannot download the token and gets an error message that would not close. The user should clear the cache to properly log out.
- Sorting on size and removing files from cart does not work in IE 10
- Annotations page not section 508 compliant
- Facets are displayed above the results table when window is small
- Table sorting icon does not include numbers

Release details are maintained in the GDC Data Portal Change Log.

0.2.15.2

GDC Data Portal Release 0.2.15-oicr2

Release Date: August 31, 2015

New Features and Changes

- Authentication, Authorization & Pop-up messages:
 - If a user tries to download a related file that is controlled from the File Entity Page, he will get a pop-up message with appropriate guidance (“Login” or “No access to the file”)
 - If a user tries to download a controlled clinical file from the Case page, he will get a pop-up message to indicate that he does not have access to the file
 - If a user authenticates to the Portal with an eRa account without dbGap authorization, he will get a warning message. Then he will have the choice to logout or to continue browsing the Portal with his eRa account. If he continues browsing the Portal with his eRa account, “My projects” filter will be hidden (temporary solution).
- If User downloads large files (from the Cart or from the File table), the Portal displays a spinner to indicate the download is in progress
- Data Download Report does not show the “Data Level” section anymore.

Bugs Fixed Since Last Release

- The add to Cart button in File Entity Page changes its display value following a click. User can then click on “Remove from Cart”
- Total Case value in the Cart matches with the number of cases associated with the files in the Cart
- When User is authenticated, “My project flag” in Case table indicates that the Cases belongs to his projects
- In Projects table, if User clicks on the count on Files, it links to the Data page - File table
- In File Entity Page, if there are no associated cases, it will display the message “No Cases Found.” instead of “No Participants Found.”

Known Issues and Workarounds

- Missing pathology files and slide images for some TCGA datasets since they are not connected to the biospecimen chain yet
- In some specific situation (complex queries using `must_not` term), the advanced search might return incorrect results
- When exporting the annotation table to a file, the create date field is not displayed in the same format than the UI
- In the advanced search, adding a trailing space to the query causes the UI to display a syntax error although the query is valid
- Checksum missing for MAGE-TAB files
- Data download statistics report is not Section 508 compliant
- User can download large files (> 1GB) from the Portal but it fails when extracting the tar.gz file. User should use the GDC Download Transfer Tool for large downloads
- Authenticated user with access to TARGET projects cannot download Target DCC files. They will get a pop-up message.
- In Case Entity Page, the “Download Clinical XML” button is not applicable to TARGET projects. The label is incorrect but the functionality is accurate.
- The advanced search suggests more fields than expected (e.g. “cases.aliqouts” should not be displayed). User should remove the field from the query and use the long field name (e.g. “cases.samples.portions.analytes.aliqouts”)
- Removing a filter after paginating table results could cause no results to be displayed. User should press Clear on the filters and start again.
- If a user has previously logged into the Portal and left a session without logging out, if he comes back to the Portal after his sessionID expires, it looks like the user is still authenticated. The user cannot download the token and gets an error message that would not close. The user should clear the cache to properly log out.

Release details are maintained in the GDC Data Portal Change Log.

0.2.15.1

GDC Data Portal Release 0.2.15-oicr1

Release Date: August 12, 2015

New Features and Changes

- Renamed all references of High Performance Download Client (HPDC) to GDC Data Transfer Tool
- Implemented more links between the Projects page and the Data page
- Improved usability and visual experience:
 - Better handling of login failure
 - User feedback when no results are available following a search
 - Warning for unsupported browsers (Internet Explorer 9 is not supported)

Bugs Fixed Since Last Release

- Optimizations for downloading large files from the browser
- Fixed various issues related to file search
- Fixed Section 508 compliance issues
- Clicking on “Total case” link in the Project List Page does not return results.

Known Issues and Workarounds

- Missing pathology files and slide images for some TCGA datasets since they are not connected to the biospecimen chain yet
- In some specific situation (complex queries using `must_not` term), the advanced search might return incorrect results

- The add to cart button in the files data view does change its display value following a click (file is correctly added to cart though)
- When exporting the annotation table to a file, the create date field is not displayed in the same format than the UI
- In the advanced search, adding a trailing space to the query causes the UI to display a syntax error although the query is valid
- Checksum missing for MAGE-TAB files
- Data download statistics report is not Section 508 compliant

Release details are maintained in the GDC Data Portal Change Log.

0.2.13

GDC Data Portal Release 0.2.13

Release Date: July 23, 2015

New Features and Changes

- Improved data portal searching. Three search mechanisms are available to the user:
 - **Facet search:** Starting with all content available on the GDC, allow users to filter down their search by clicking on elements on the left of the screen. This feature is available in **Projects** and **Data** page. Range support was also added to facets in this release
 - **Advanced Search:** Starting with all content available on the GDC, allow users to build a custom and complex query using all of GDC capabilities (any field with the parameter of their choice).
 - **Quick Search:** When looking for specific portal element, allow users to launch the **Quick Search** by clicking the “?” or “CTRL+SPACE” and find high-level informations of some entities.
- Updated styling to align with NCI new visual identify
- Created a pie chart widget allowing user to easily switch between a pie chart and a table view.
- Improved Usability and visual experience:
 - Added tooltips to various sections of the portal
 - Added a range facet with barchart
 - Add more charts (summary, cart)

Bugs Fixed Since Last Release

- Hooked-up reports to real data
- Fixed various issues on GQL (Advanced Search)
- Table export to export appropriate columns
- Allow users to sort project list table

Known Issues and Workarounds

- Checksum missing for MAGE-TAB files

Release details are maintained in the GDC Data Portal Change Log.

0.1.10

GDC Data Portal Release 0.1.10

Release Date: March 18, 2015

New Features and Changes

- Authentication and Authorization
 - Allow users to authenticate to the portal using their to ERA Commons credentials
 - Allow users to narrow down their search based on their own projects
 - Allow users to download a token link (for use with the API) in the portal
 - Allow users to download protected data (if they have permissions to do so)
- Reports
 - Allow users to view a data download statistics report
- Improve usability and visual experience:
 - Allow users to view projects data using a new type of graph (“github” style)
 - Implement more features into the table widget (sort per column, hide/show columns, re-arrange columns, export to JSON, XML, CSV, TSV, maintain user preferences)
 - Display UI and API version at the bottom of the page
- Search
 - Allow users to select multiple terms in facets (OR operand)
 - Improvements on advance search with auto-loading of possible fields
- Cart
 - Allow users to view more details through additional pie charts
 - Allow users to download a manifest from the cart
 - Improved the mechanism of adding data to cart in various sections of the portal.
- Updated style/theme to match GDC Website
- Display a NCI Warning banner to inform users about GDC policy

Bugs Fixed Since Last Release

- Improvements in 508 compliance

Known Issues and Workarounds

- TARGET data is currently not available
- Data:
 - Absence of “real” download statistics for the data download statistics report
 - Missing checksum for magetab files in file entity page
- Project list table not sortable

Release details are maintained in the GDC Data Portal Change Log.

0.1.8

GDC Data Portal Release 0.1.8

Release Date: January 22, 2015

New Features and Changes

- Allow users to perform a project search and obtain a list of projects (Project Search and List Pages)
- Allow users to retrieve project details (Project Entity Page)
- Allow users to perform an annotations search and obtain a list of annotations (Annotations Search and List Pages)
- Allow users to retrieve entity details (Annotation Entity Page)
- Implement a basic search feature (Basic Search - Facets))
- Implement an advance search feature (Advanced Search - Query Language)
- Allow users to retrieve file details (File Entity Page)
- Allow users to retrieve participant details (Participant Entity Page)
- Allow users to view and add files to a cart (Cart)
- Allow users to view reports (Reports - Data Download Statistics)
- Allow users to export tables of search results (Export Tables)
- Allow users to download files (Download)
- Allow users to authenticate using eRA Commons (Authentication)

Bugs Fixed Since Last Release

- Initial Release - Not Applicable

Known Issues and Workarounds

- TARGET data is currently not available

Release details are maintained in the GDC Data Portal Change Log.