caGWAS Data Portal 1.1.0

User's Guide

October 8, 2009

CONTENTS

About This Guide

This section introduces you to the *caGWAS Data Portal 1.1.0 User's Guide*. It includes the following topics:

Purpose

This guide provides an overview of the caGWAS data portal. It explains how to browse data in the portal, create or request a report, and download bulk data.

Topics Covered

If you are new to the caGWAS data portal, read this brief overview, which explains what you will find in each chapter and appendix.

- *Chapter 1, About the caGWAS Data Portal,* on page 3 introduces the caGWAS data portal and explains how to log in and access the online help.
- *Chapter 2, Browsing caGWAS Data,* on page 7 describes how to query and browse SNP Association Findings, SNP Population Frequency, and SNP Subjects data, as well as how to create reports of that data.
- *Chapter 3, Downloading Bulk Data,* on page 25 explains how to download large data sets.
- *caGWAS Glossary* on page 27 is a glossary of terms related to the caGWAS data portal.

Text Conventions Used

This section explains conventions used in this guide. The various typefaces represent interface components, keyboard shortcuts, toolbar buttons, dialog box options, and text that you type.

Convention	***Description***	***Example***
Bold	Highlights names of option buttons, check boxes, drop-down menus, menu commands, command buttons, or icons.	Click **Search**.
URL	Indicates a Web address.	http://domain.com
text in SMALL CAPS	Indicates a keyboard shortcut.	Press ENTER.
text in SMALL CAPS + text in SMALL CAPS	Indicates keys that are pressed simultaneously.	Press SHIFT + CTRL.
Italics	Highlights references to other documents, sections, figures, and tables.	See *Figure 4.5*.
Italic boldface monospaced type	Represents text that you type.	In the **New Subset** text box, enter ***`Proprietary Proteins.`***
Note:	Highlights information of particular importance	**Note:** This concept is used throughout the document.
{ }	Surrounds replaceable items.	Replace {last name, first name} with the Principal Investigator's name.

CHAPTER

1

ABOUT THE CAGWAS DATA PORTAL

The Cancer Genome-Wide Association Studies (caGWAS) data portal allows researchers to integrate, query, report, and analyze significant associations between genetic variations and disease, drug response or other clinical outcomes. New breakthroughs in SNP array technologies make it possible to genotype hundreds of thousands of single nucleotide polymorphisms (SNPs) simultaneously, enabling whole genome association studies. Within the Clinical Genomic Object Model (CGOM), the caIntegrator team created a domain model for Whole Genome Association Study Analysis. CGOM-caGWAS is a semantically annotated domain model that captures associations between Study, Study Participant, Disease, SNP Association Analysis, SNP Population Frequency and SNP annotations.

This chapter includes the following topics:

Logging In to the caGWAS Data Portal

Logging in to the caGWAS data portal allows you to search subject, genotype, association finding, and population frequency data.

To log in to caGWAS

1. Enter your **Username** and **Password**.

   Note: If you forgot your password, go to https://password.nci.nih.gov to reset it. Click Forgotten Password for tips on choosing a new password.

2. Click **Submit**. The caGWAS Data Portal home page appears.

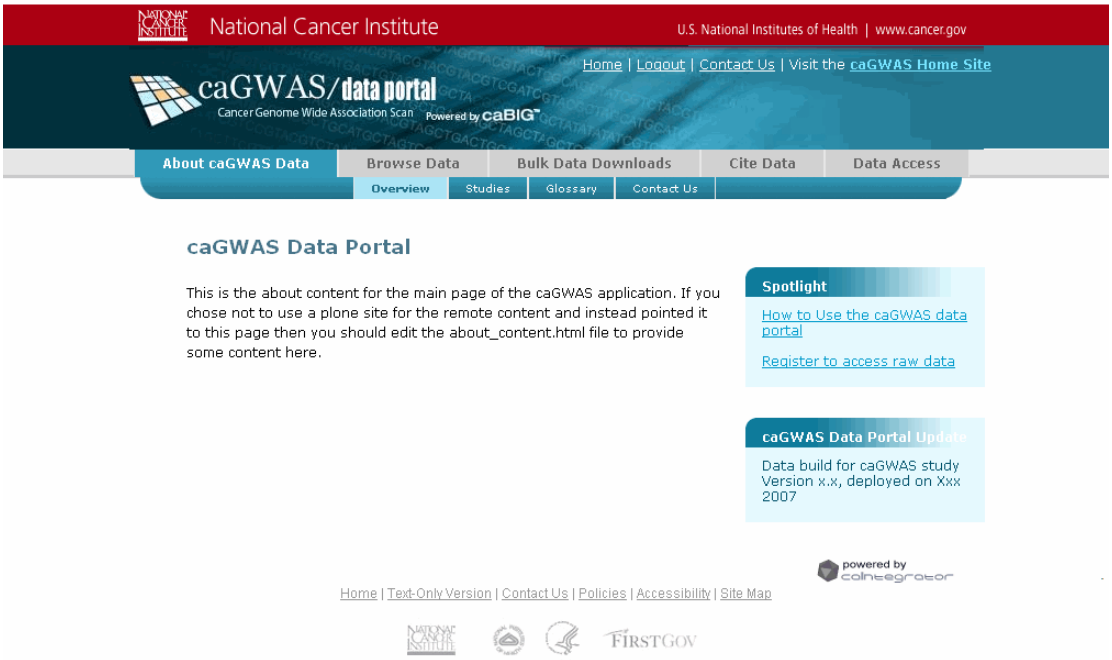


Figure 1.1 caGWAS Data Portal Home Page

If you are not a registered user, you can apply for an account.

To apply for an account to access the caGWAS data portal

1. At the top of any caGWAS page, click the **Login** link. The page where you can either log in or register for access appears.

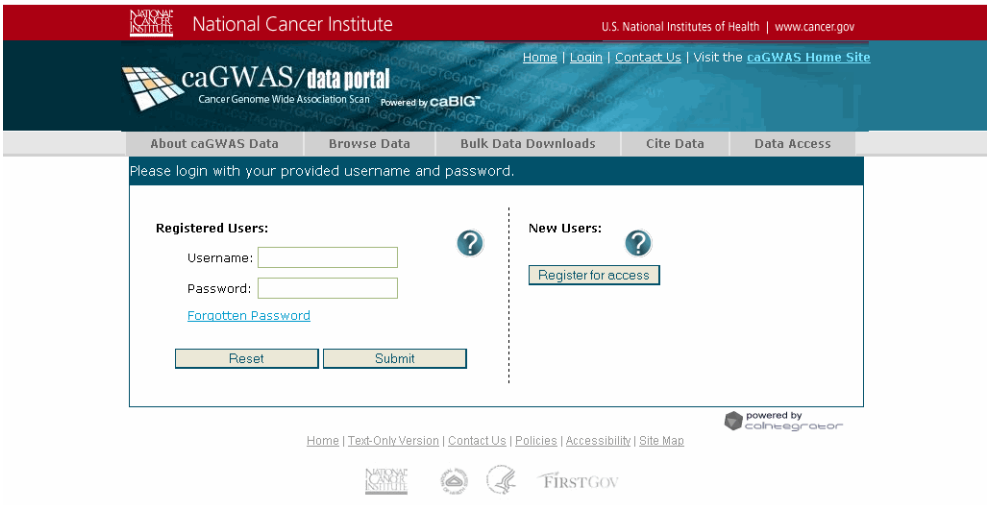


Figure 1.2 caGWAS Login and Register for Access Page

2. Click the **Register for Access** button. If you were to customize this page, instructions about how to become a registered user would appear.

Using caGWAS Online Help

The online help explains how to use all of the features of the caGWAS data portal.

To access online help in the caGWAS data portal, use either of the following methods:

- Click **How to Use the caGWAS data portal** in the box in the right-hand column of the home page to access all help pages.
- Click on any page to get help for the information or tasks on that page.

See the following tips for locating topics of interest, navigating the help, and printing a topic.

Click this	***To do this***
	Highlight your current topic in the table of contents.
	Print your current topic.
	When available at the bottom of a topic or in the upper-right hand corner, see topics related to the current topic.
Breadcrumb trail link	Open that help topic. The breadcrumb trail at the top of the page shows the location of the current help topic relative to neighboring topics. An example is: Browsing caGWAS Data > Searching SNP Data by Association Finding
Index tab	View the help index.
Search tab	Search for one or more words by entering a query in the text field.
Favorites tab	Create a list of help topics you want to return to frequently by navigating to the topic, clicking the **Favorites** tab at the top of the table of contents pane, and then clicking **Add**.
Divider line	To expand or narrow the width of the table of contents or topic pane, click and drag the divider between them.

Table 1.1 Locating Topics of Interest in Online Help

The following icons represent unique information in a help topic.

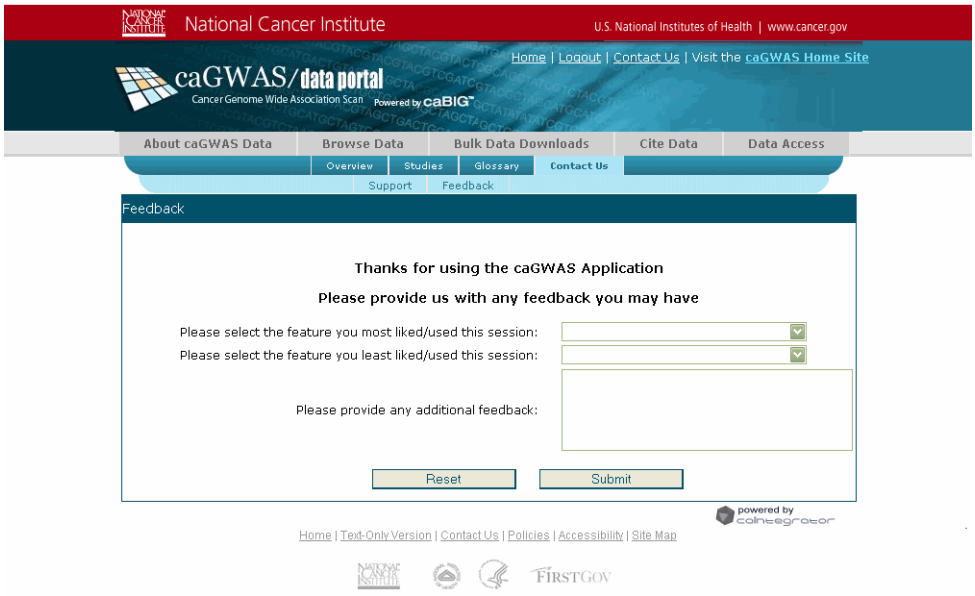
	Indicates useful information
	Indicates practical advice or additional points relating to the text you just read
	Alerts you to critical steps to take or advice to follow

Providing Feedback

Please help us improve the caGWAS data portal by providing your feedback.

To provide feedback to the caGWAS team

1. Navigate to **About caGWAS Data** > **Contact Us** > **Feedback**. The Feedback page appears.



2. Select caGWAS features from either of the two lists, as you prefer.
3. In the text box, enter any other feedback you have.
4. Click **Submit**.

   To clear your feedback and start over, click **Reset**.

The NCI does not track any information through this feedback form that can identify you. However, if you would like to be contacted, please provide your contact information in the text box.

CHAPTER

2

BROWSING CAGWAS DATA

You can search the caGWAS database for specific genes or regions of interest. You can browse by:

- **Association finding**, to identify regions of chromosomes associated with cancer
- **Population frequency**, to determine the SNP frequency by population type
- **Subjects**, to retrieve data specific to patients and controls (available to registered users only)

This chapter includes the following topics.

Searching SNP Data by Association Finding

This type of search identifies regions of chromosomes associated with cancer. For each study in the caGWAS data portal, you may refine your search by selecting analysis categories uniquely relevant to the study.

To search by association finding

1. Click the **Browse Data** menu. The Browse Data page appears.

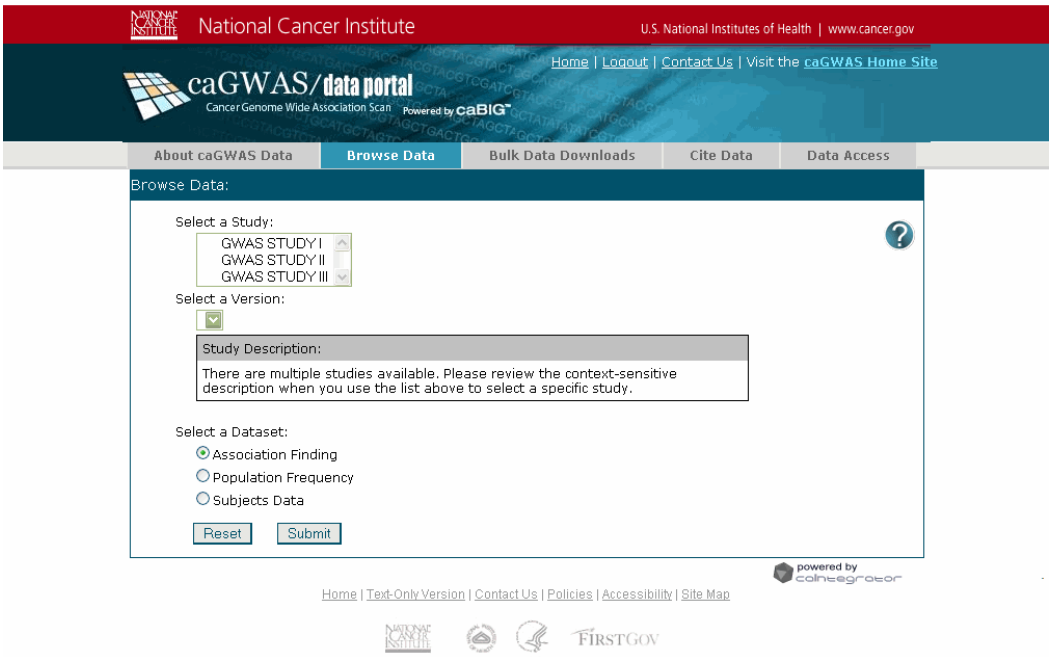


Figure 2.1 caGWAS Browse Data page

2. Select a study and a version. If you do not select a version, your search returns the most recent study data.
3. Click the **Association Finding** dataset option.
4. Click **Submit**. The Search Association Findings page appears.

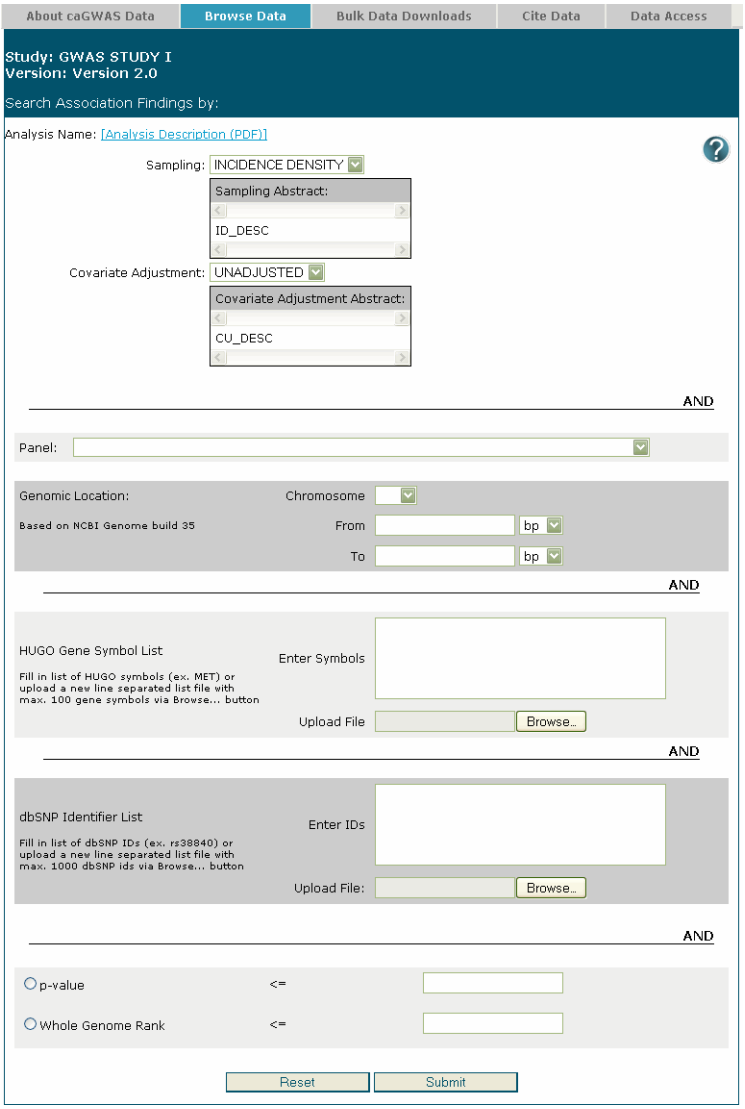


Figure 2.2 Search Association Findings page

Note: All of the following search criteria are optional. The more you refine your search by adding search criteria, the fewer results you receive. On the Browse Data page you must select at least one search criterion to filter your results. If you want to see the entire dataset, select the Bulk Data Download menu instead. For more information, see *Downloading Bulk Data* on page 25.

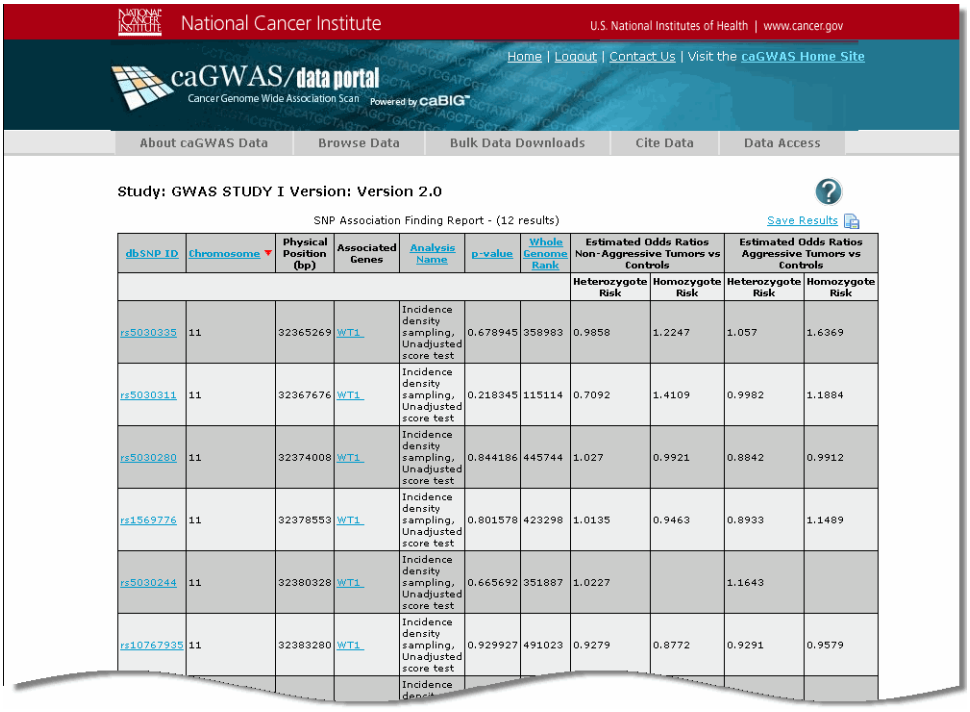
5. In the Analysis Name section, select an analysis category from each of the available drop-down lists. After you select a category, a definition of that category appears below the list. For more information about the analysis categories, click the **Analysis Description (PDF)** link at the top of the page.
6. To further refine your search, specify any combination of the following criteria.

Table 2.1 Search criteria available for association finding search

Search Criteria	Explanation
Panel	Select a SNP panel.
Genomic Location	To search by genomic location, select a chromosome number from the drop-down menu. Specify a specific chromosome region in the **From** and **To** boxes.
HUGO Gene Symbols	To search by gene symbol, enter one or more HUGO gene symbols (for example, MET) in the text block, or, to upload a list of up to 100 gene symbols, click **Browse** to select the file.
dbSNP Identifiers	To search by dbSNP ID, enter one or more dbSNP IDs (for example, rs38840), or, to upload a list of up to 1000 dbSNP IDs, click **Browse**.
p-value	To search by p-value, select **p-value** to enter a value between 0 to 1.
Whole Genome Rank	To search by rank of as so cation, select **Whole Genome Rank** and enter an integer that represents a cut-off rank.

Note: To clear your selections and start a new search, click **Reset**.

7. Click **Submit.** The SNP Association Findings Report appears.



dbSNP ID	Chromosome	Physical Position (bp)	Associated Genes	Analysis Name	p-value	Whole Genome Rank	Estimated Odds Ratios Non-Aggressive Tumors vs Controls		Estimated Odds Ratios Aggressive Tumors vs Controls	
							Heterozygote Risk	Homozygote Risk	Heterozygote Risk	Homozygote Risk
rs5030335	11	32365269	WT1	Incidence density sampling, Unadjusted score test	0.678945	358983	0.9858	1.2247	1.057	1.6369
rs5030311	11	32367676	WT1	Incidence density sampling, Unadjusted score test	0.218345	115114	0.7092	1.4109	0.9982	1.1884
rs5030280	11	32374008	WT1	Incidence density sampling, Unadjusted score test	0.844186	445744	1.027	0.9921	0.8842	0.9912
rs1569776	11	32378553	WT1	Incidence density sampling, Unadjusted score test	0.801578	423298	1.0135	0.9463	0.8933	1.1489
rs5030244	11	32380328	WT1	Incidence density sampling, Unadjusted score test	0.665692	351887	1.0227		1.1643	
rs10767935	11	32383280	WT1	Incidence density sampling, Unadjusted score test	0.929927	491023	0.9279	0.8772	0.9291	0.9579
				Incidence						

Figure 2.3 Sample SNP Association Findings report

If your query results in more than 500 records, caGWAS generates a report and sends it to you by email. For more information, see *Downloading Bulk Data* on page 25.

See *Understanding the SNP Association Findings Report* on page 11 for more information.

Understanding the SNP Association Findings Report

The SNP Association Findings report provides information on SNPs analyzed in a genome-wide analysis, indicating the degree of association found between SNPs and phenotypes.

Explanations of the columns in the SNP Association Findings report appear in the following table.

Table 2.2 Explanation of data in SNP Association Findings report

Column Number	Column	Explanation
1	*dbSNP ID*	SNP identifier, see http://www.ncbi.nlm.nih.gov/SNP
2	*Chromosome*	Chromosome number Click the column name to sort the report by chromosome and physical location.
3	*Physical Position (bp)*	Position of the SNP on the chromosome in base pairs
4	*Associated Genes*	Genes within 15,000 base pairs upstream and 10,000 base pairs downstream of the SNP
5	Analysis Name	• Incidence density sampling, Unadjusted score test or • Incidence density sampling, Adjusted score test
6	*p-value*	Significance value of the analysis - a lower value means the result is more significant
7	*Whole Genome Rank*	Rank of the SNP's p-value compared to all other SNPs in the genome scan (higher rank number indicates lower significance)
8	Estimated Odds Ratios Non-Aggressive Tumors vs. Controls	Estimated odds ratios for non-aggressive tumors vs. controls for heterozygote and homozygote risk
9	Estimated Odds Ratios Aggressive Tumors vs. Controls	Estimated odds ratios for aggressive tumors vs. controls for heterozygote and homozygote risk

Notes:

- The SNPs are in order of p-value for the unadjusted analysis.

- A SNP that is not close to a gene (as defined in column 4 above) will not have a value for "Associated Genes".
- There are two rows for each SNP - one for the adjusted score, and one for the unadjusted score.
- Description of analyses
  - Unadjusted: A 3-by-m contingency table of genotypes by phenotypes was constructed.
  - Adjusted: The m phenotypes were regressed on indicator variables for genotype effects, age group at randomization (4 groups), region of recruitment (9 non-reference regions), and an indicator variable for cases diagnosed within one year of entry to the trial, and 3 sets of eigenvectors from the most significant principal components.

You can click a column name link to sort the whole report by that column. Sorting order toggles between ascending order (low to high) and descending order (high to low) by repeated clicks. An arrow image indicates the sorted order of the column as  or  .

To save the data to a tab-delimited text file, click **Save Results**. Your browser prompts you to open or save the file. For convenient data viewing, save the file and open it in a spreadsheet application.

For more information, see the following topics.

- *Searching SNP Data by Association Finding* on page 7

Searching SNP Data by Population Frequency

This type of search finds the frequency of SNPs in certain population groups.

For detailed information about these population groups, begin to browse data by association finding and click the **Analysis Description (PDF)** link in the Analysis Name text block. For information on how to browse data by association finding, see *Searching SNP Data by Association Finding* on page 7.

To search by population frequency

1. Click the **Browse Data** menu. The Browse Data page appears.

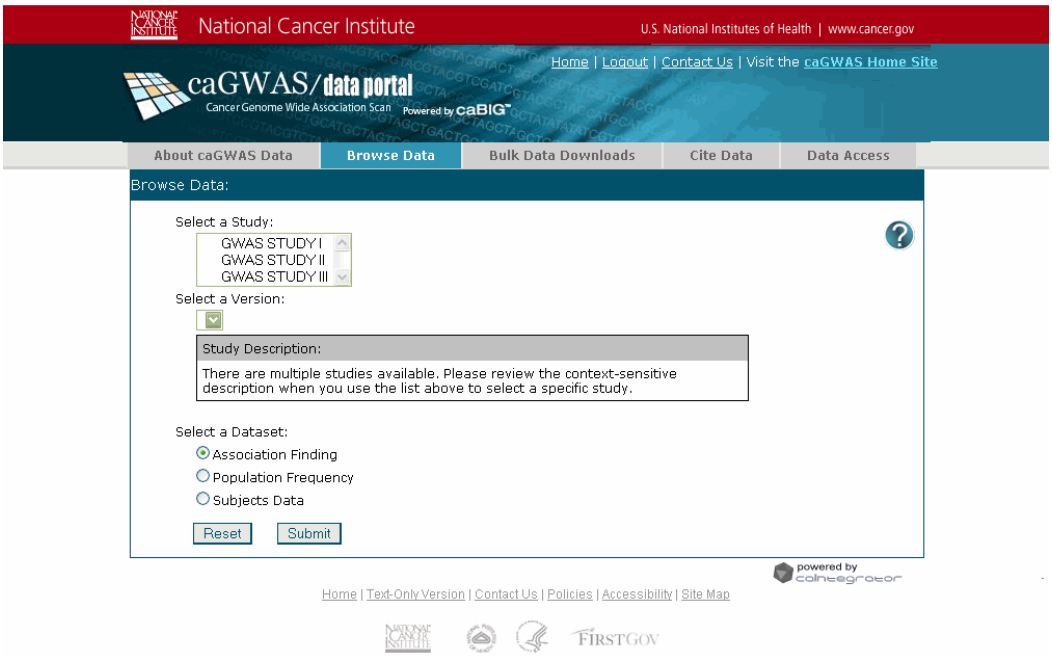


Figure 2.4 caGWAS Browse Data page

2. Select a study and a version. If you do not select a version, your search returns the most recent study data.
3. Click the **Population Frequency** dataset option.

4. Click **Submit**. The query page for searching SNP Population Frequency data appears.

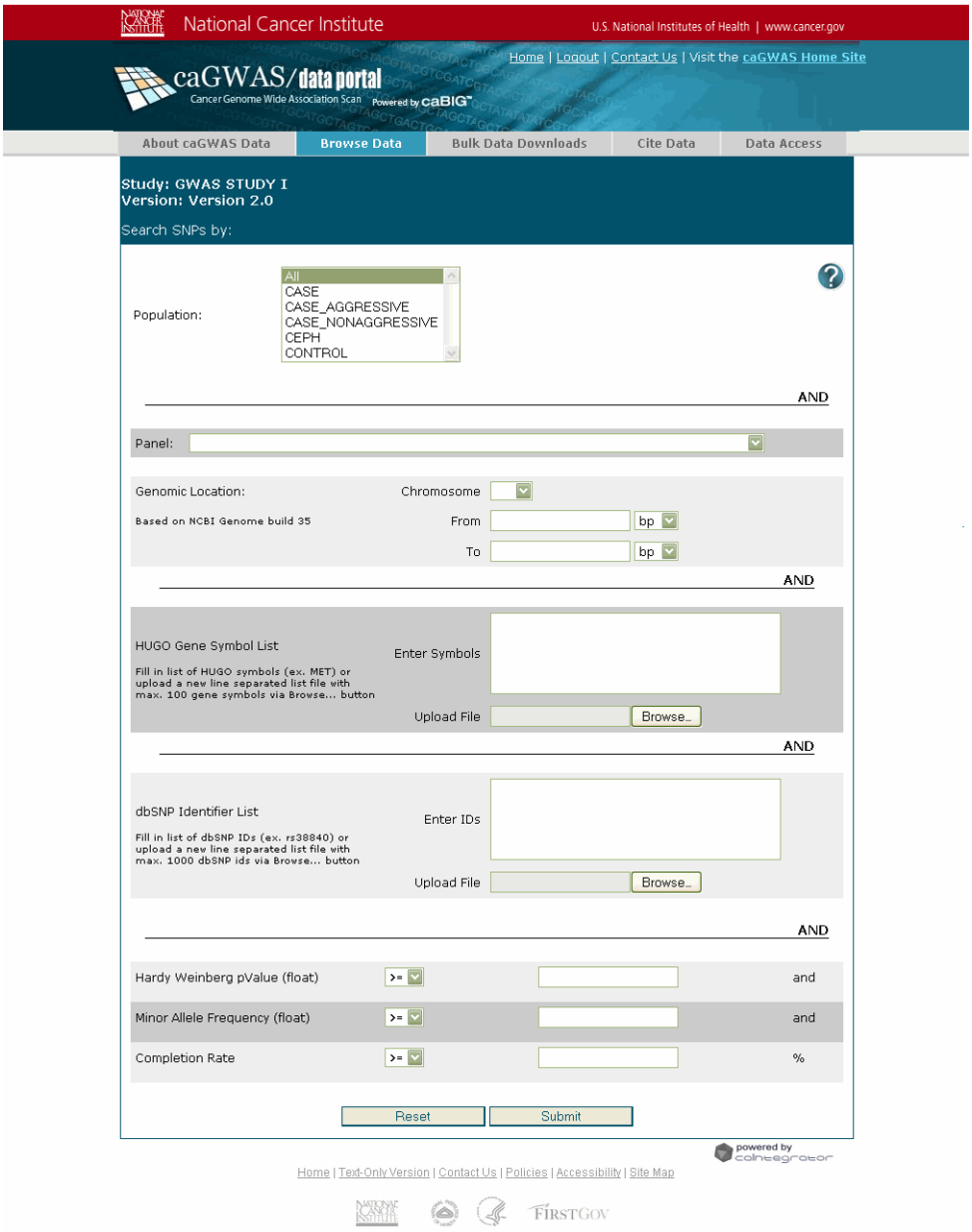


Figure 2.5 Search SNP Population Frequency page

5. Select the Population type from the drop-down menu. To select more than one group, press CTRL while clicking each selection. If you do not select a population, your search returns data from all populations.
6. To further refine your search, specify any combination of the following criteria.

Note: All of the following search criteria are optional. The more you refine your search by adding search criteria, the fewer results you receive. On the Browse Data page you must select at least one search criterion to filter your results. If you want to see the entire dataset, select the Bulk Data Download menu instead. For more information, see *Downloading Bulk Data* on page 25.

Table 2.3 Search criteria available for population frequency search

Column Number	Search Criteria	Explanation
1	*Panel*	Select a SNP panel.
2	*Genomic Location*	To search by genomic location, select a chromosome number from the drop-down menu. Specify a specific chromosome region in the **From** and **To** boxes.
3	*HUGO Gene Symbols*	To search by gene symbol, enter one or more HUGO gene symbols (for example, MET) in the text block, or, to upload a list of up to 100 gene symbols, click **Browse** to select the file.
4	*dbSNP Identifiers*	To search by dbSNP ID, enter one or more dbSNP IDs (for example, rs38840), or, to upload a list of up to 1000 dbSNP IDs, click **Browse**.
5	*Hardy Weinberg pValue*	Select greater than [>=] or less than [<=] from the drop-down menu. Enter a number in the text block.
6	*Minor Allele Frequency*	Select greater than [>=] or less than [<=] from the drop-down menu. Enter a number in the text block.
7	*Completion Rate*	Select greater than [>=] or less than [<=] from the drop-down menu. Enter a number in the text block.

Note: To clear your selections and start a new search, click **Reset**.

7. Click **Submit**. The SNP Population Frequency Report appears.

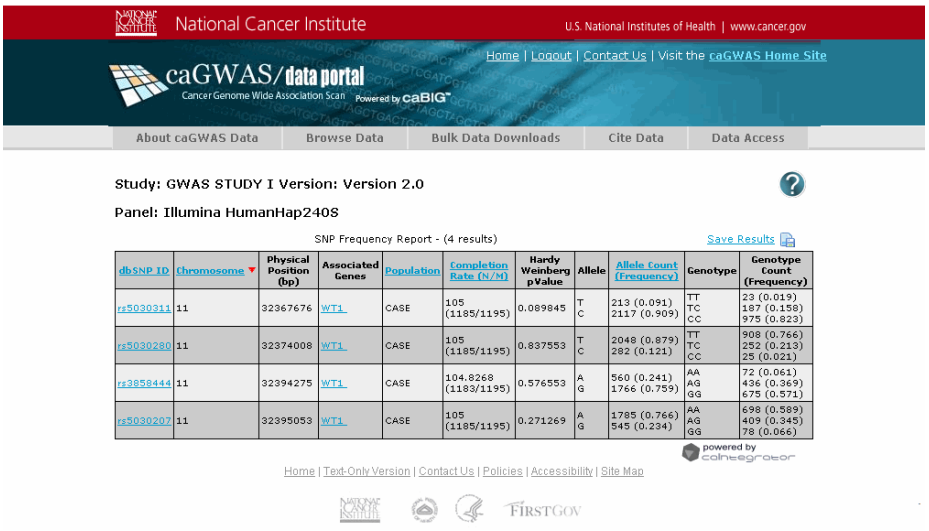


Figure 2.6 Sample of a SNP Population Frequency report

If your query results in more than 500 records, caGWAS generates a report and sends it to you by email. For more information, see *Downloading Bulk Data* on page 25.

See *Understanding the SNP Population Frequency Report* on page 17 for more information.

Understanding the SNP Population Frequency Report

The SNP Population Frequency report provides information on SNPs analyzed in a caGWAS genome-wide analysis, indicating the allelic and genotypic frequencies.

Following are explanations of the columns in the SNP Population Frequency report. To see column definitions in a popup window, hover your mouse over words that have a maroon, dotted link.

Table 2.4 Explanation of data in SNP Population Frequency report

Column Number	***Column***	***Explanation***
1	*dbSNP ID*	SNP identifier, see http://www.ncbi.nlm.nih.gov/SNP Click link for details of SNP.
2	Chromosome	Chromosome number Click the column name to sort the report by chromosome and physical location.
3	Physical position (bp)	Position of the SNP on the chromosome in base pairs
4	Associated Genes	Genes within 15,000 base pairs upstream and 10,000 base pairs downstream of the SNP Click link for details of gene.
5	Population	Population group selected in search, which can be any of the following: • CASE • CASE_AGGRESSIVE • CASE_NONAGGRESSIVE • CEPH • CONTROL
6	*Completion rate (N/M)*	N is the count of the non-missing genotypes, M is the total number of attempted genotypes
7	*Hardy Weinberg p-value*	Exact significance test of the Hardy-Weinberg assumption, the probability that deviation from Hardy Weinberg equilibrium could be explained by chance
8	*Allele*	The SNP's alleles, e.g. "T|C", the order of alleles is arbitrary
9	Allele Count (frequency)	Count and frequency of each allele; for example, "1860(0.842)|348(0.158)" indicates 1860 occurrences of the T allele, frequency .842, 348 occurrences of T allele, frequency.158 Click the column name to sort the report by the Minor Allele Frequency (MAF).
10	*Genotype*	The SNP's genotypes, e.g. "TT|TC|CC", the order is based on the allele order (see column 8)

Table 2.4 Explanation of data in SNP Population Frequency report (Continued)

Column Number	Column	Explanation
11	Genotype Count (frequency)	Count and frequency of each genotype, e.g. "1009(0.914)|93(0.084)|2(0.002)" indicates 1009 occurrences of the TT genotype, frequency .914, 93 occurrences of the TC genotype, frequency .084, 2 occurrences of the CC genotype, frequency .002

Notes:

- The SNPs are in order of dbSNP ID (rs number).
- A SNP that is not close to a gene (as defined in column 4 above) will not have a value for "Associated Genes".

You can click a column name link to sort the whole report by that column. Sorting order toggles between ascending order (low to high) and descending order (high to low) by repeated clicks. An arrow image indicates the sorted order of the column as  or  .

To save the data to a tab-delimited text file, click **Save Results**. Your browser prompts you to open or save the file. For convenient data viewing, save the file and open it in a spreadsheet application.

For more information, see the following topic:

- *Searching SNP Data by Population Frequency* on page 12

Searching SNP Subject Data

Only registered users may access data about subjects and controls.

To search by subject data

1. Click the **Browse Data** menu. The Browse Data page appears.

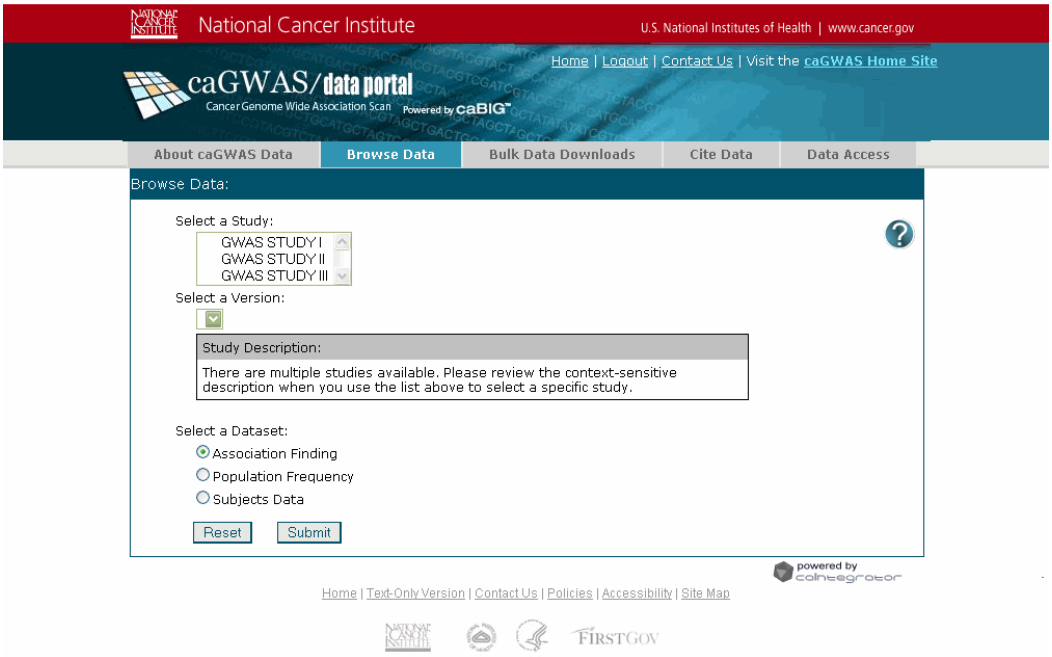


Figure 2.7 caGWAS Browse Data Page

2. Select a study and a version. If you do not select a version, your search returns the most recent study data.
3. Click the **Subject Data** dataset option.

4. Click **Submit**. The Search Subject Data page appears.

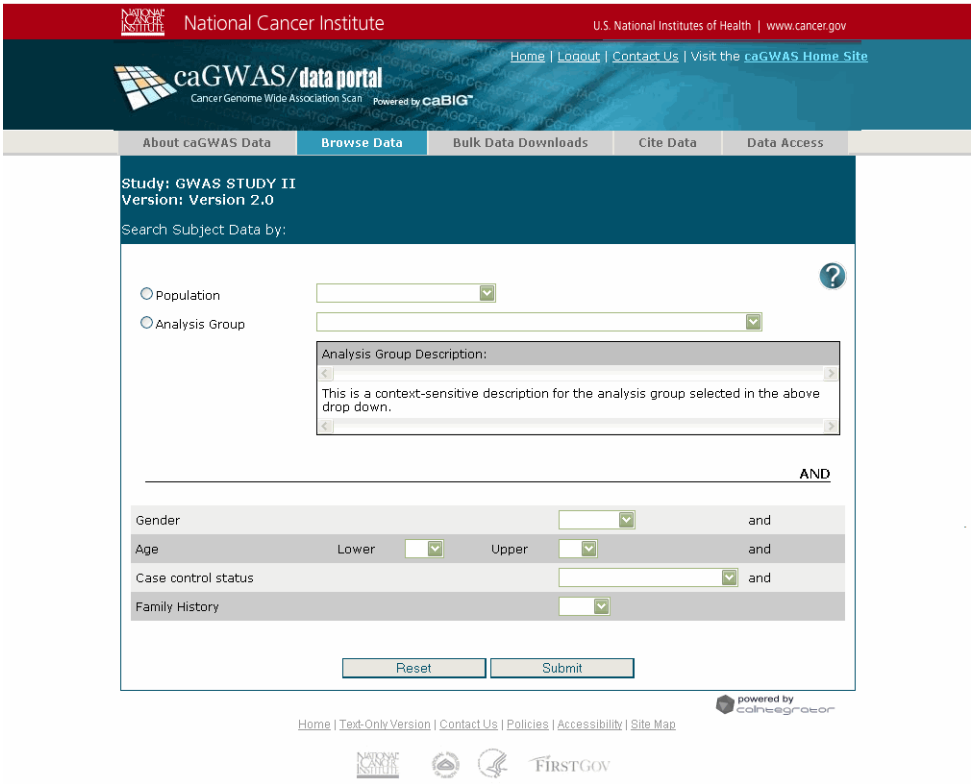


Figure 2.8 Search Subject Data page

5. To restrict your SNP subject data search to a specific population or analysis group, select one from the drop-down menus. An abstract of the analysis appears in the Analysis Group Description text block.

   If you do not select a population or an analysis group, the search returns data from all populations and analysis groups.

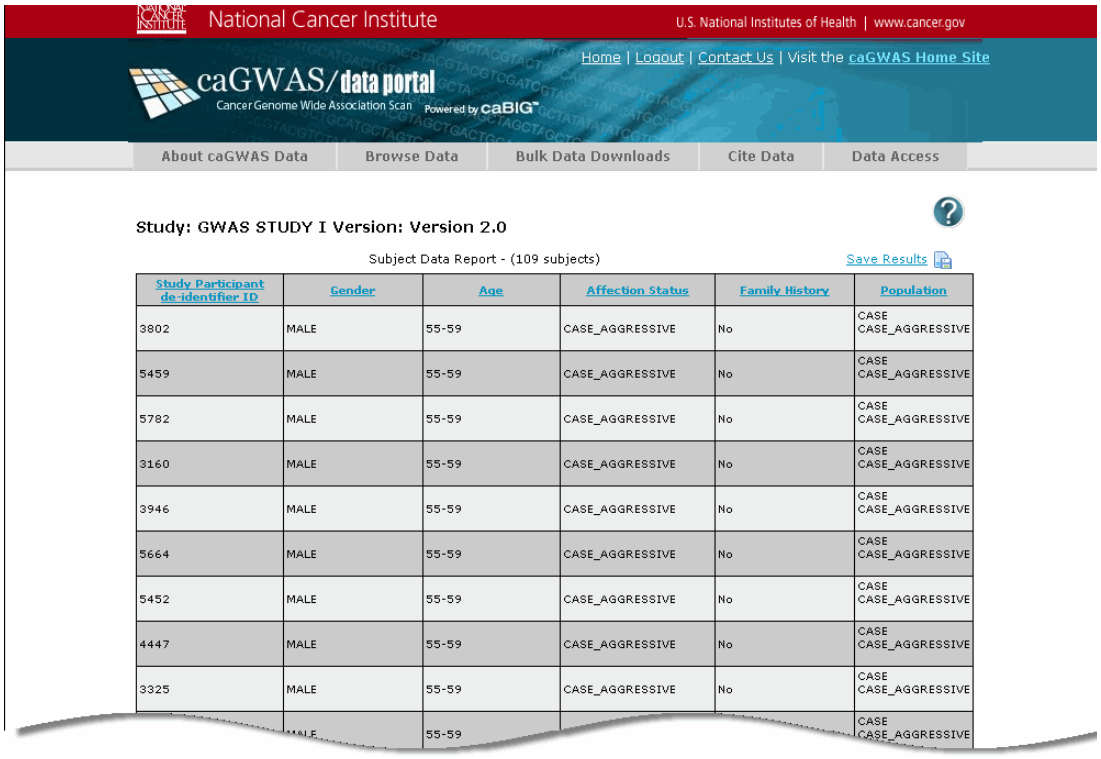
6. You can further refine your search by specifying additional information using the search criteria described in the following table.

   Table 2.5 Additional search criteria for subject data search

Search Criteria	***Explanation***
Gender	Select gender from the drop-down menu
Age	Select upper and lower range from the drop-down menus
Case control status	Study participant association with disease status
Family History	Select **yes** or **no** from the drop-down menu

Note: To clear your selections and start a new search, click **Reset**.

7. Click **Submit**. The SNP Subject Data Report appears.



Study Participant de-identifier ID	Gender	Age	Affection Status	Family History	Population
3802	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
5459	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
5782	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
3160	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
3946	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
5664	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
5452	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
4447	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
3325	MALE	55-59	CASE_AGGRESSIVE	No	CASE CASE_AGGRESSIVE
		55-59			CASE CASE_AGGRESSIVE

Figure 2.9 Sample of a SNP Subject Data report

If your query results in more than 500 records, caGWAS generates a report and sends it to you by email. For more information, see *Downloading Bulk Data* on page 25.

See *Understanding the SNP Subject Data Report* on page 21 for more information.

Understanding the SNP Subject Data Report

The SNP Subject Data report provides information on the subjects in a selected caGWAS study.

Explanations of the columns in the SNP Subject Data report appear in the following table.

Table 2.6 Explanation of data in SNP Subject Data report

Column Number	***Column***	***Explanation***
1	Study participant de-identifier ID	Study Participant de-identified ID
2	Gender	Gender - MALE or FEMALE
3	Age	Age range of the patient when entered into the study

Table 2.6 Explanation of data in SNP Subject Data report (Continued)

Column Number	Column	Explanation
4	Affected Status	Overall status for each subject in the study - can be any of the following: • CASE_AGGRESSIVE • CASE_NONAGGRESSIVE • CONTROL
5	Family History	Yes or No
6	Population	Group of subjects for computing SNP frequencies - can be any of the following: • CASE_AGGRESSIVE • CASE_NONAGGRESSIVE • CEPH • CONTROL

Note: For CEPH population subjects, the affected status is blank.

You can click a column name link to sort the whole report by that column. Sorting order toggles between ascending order (low to high) and descending order (high to low) by repeated clicks. An arrow image indicates the sorted order of the column as ▼ or ▲.

To save the data to a tab-delimited text file, click **Save Results**. Your browser prompts you to open or save the file. For convenient data viewing, save the file and open it in a spreadsheet application.

For more information, see the following topic:

- *Searching SNP Subject Data* on page 19

Requesting Data for More than 500 Records

If your SNP Association Finding, SNP Population Frequency, SNP Subjects Data, or Genotypes Data report exceeds 500 records, you do not receive that report immediately. caGWAS prepares a report and emails it to you.

caGWAS informs you of the size of the report you requested by displaying the following window.

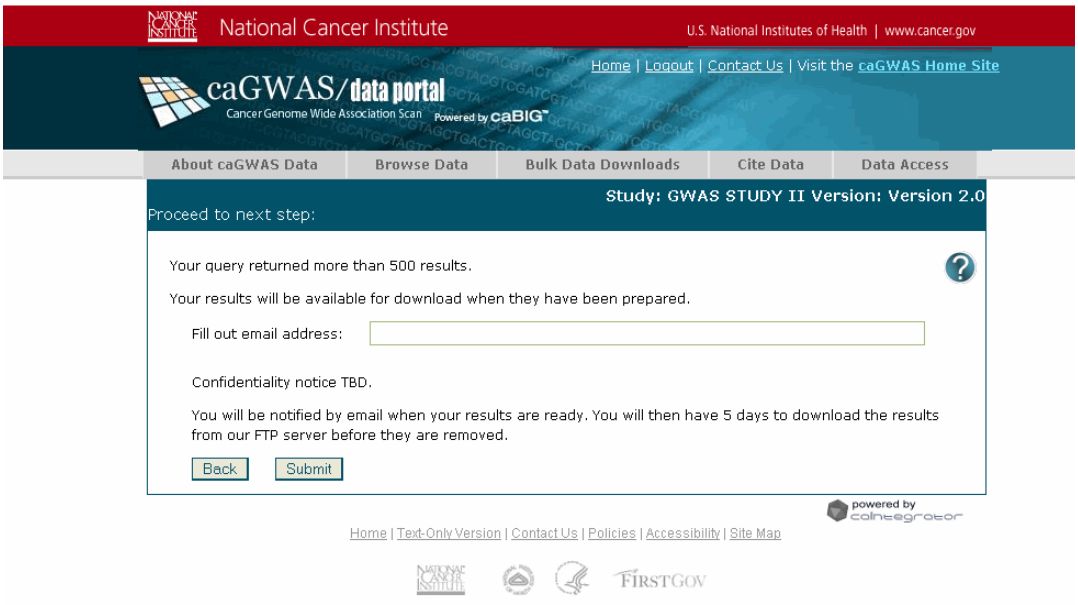


Figure 2.10 Results page when query returns more than 500 records

In the page that appears after you submit your query, enter your email address where requested and then click **Submit**. You will receive an email when your report is available to download. You will have five days to download the report.

CHAPTER

3

DOWNLOADING BULK DATA

Your caGWAS data portal access privileges may allow you to download bulk data.

Downloading Bulk Data

Registered users may download bulk data associated with the study which they have permission to access.

Registered users can download bulk data for:

- Association Finding—A single tab delimited file for all association analysis results
- Population Frequency—A single tab delimited file for all population frequency results and separate files of population frequency results by population
- Subjects Data—A single tab delimited file for all study participant
- Genotype Data—A single tab delimited file for all genotype data, and separate files of genotype data by chromosome, by population, and by both chromosome and population

To download bulk data

1. Log in to the caGWAS data portal by clicking the **Login** link at the top of the page.
2. Click the **Bulk Data Downloads** menu. The Bulk Data Download page appears.

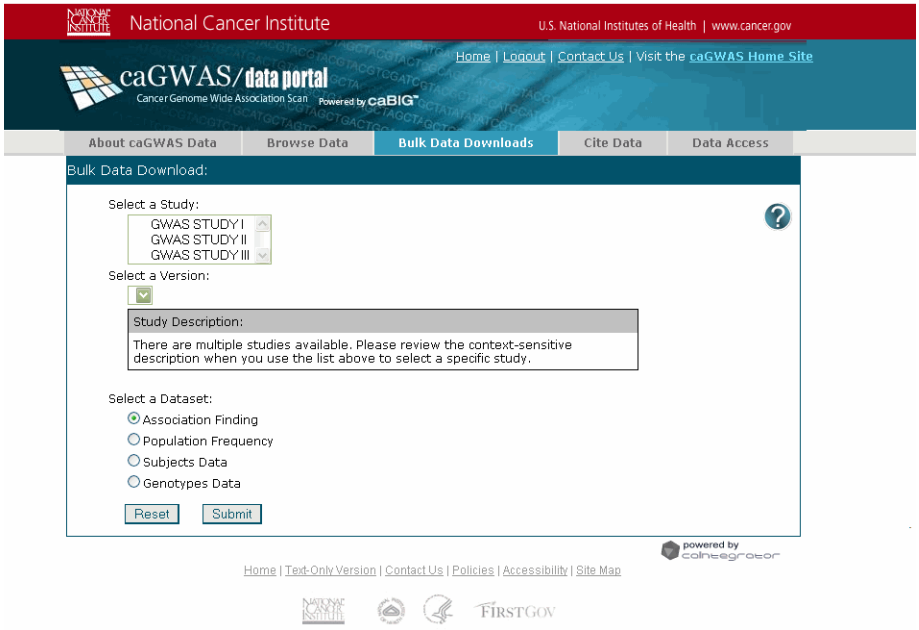


Figure 3.1 Bulk Data Download page

3. Select a study and a version.
4. Select a dataset from the following options:
   - Association Finding
   - Population Frequency
   - Subjects Data
   - Genotypes Data
5. Click **Submit**. A web page appears where you can download files matching your search criteria.

   If you are not a registered user, apply for an account by clicking the **Login** link on the caGWAS data portal home page and then clicking the **Register for Access** button. A page of instructions for becoming a registered user appears.

caGWAS Glossary

Acronyms, objects, tools and other terms referred to in this online help are described in this glossary.

Term	*Definition*
allele	Any one of a number of viable DNA codings occupying a given locus (position) on a chromosome
caGWAS	Cancer Genome-Wide Association Study
case-control study	A study including subjects who already have a condition, and those who don't have the condition, to determine if there are characteristics of the affected subjects that differ from the unaffected subjects
CBIIT	Center for Biomedical Informatics and Information Technology
CGEMS	Cancer Genetic Markers of Susceptibility
completion rate	For a set of genotype data (either by SNP or for a single individual), the percentage of genotypes completed successful compared to genotypes attempted
confidence interval (CI)	A range around a measurement that conveys how precise the measurement is
dbSNP identifier	The identifier for a cluster of polymorphisms in dbSNP, NCBI's central repository for single base nucleotide substitutions (SNPs) and short deletion and insertion polymorphisms (aka "rs number") - see http://www.ncbi.nlm.nih.gov/SNP
genome	The complete sequence of DNA contained in an organism or a cell, including both the chromosomes within the nucleus and the DNA in mitochondria
Genome-wide association study (GWAS)	An approach that involves rapidly scanning markers across a person's genome to find SNPs associated with a particular condition
genomic location	The physical location of a feature (e.g. gene, exon, SNP) on a genome or chromosome
genotype	The genetic makeup encoded in an individual's DNA. When related to SNPs, the genotype refers to the nucleotides at the SNP locus on the two DNA strands of the sample.

Term	Definition
Hardy Weinberg equilibrium	The Hardy-Weinberg principle (HWP) states that, under certain conditions, after one generation of random mating, the genotype frequencies at a single gene locus will become fixed at a particular equilibrium value. It also specifies that those equilibrium frequencies can be represented as a simple function of the allele frequencies at that locus.
HUGO gene symbol	A gene symbol approved by and included in the HGNC (HUGO Gene Nomenclature Committee database) - see http://www.gene.ucl.ac.uk/nomenclature
minor allele frequency	The frequency of chromosomes in the population carrying the less common variant of SNP
NCBI	National Center for Biotechnology Information, see http://www.ncbi.nlm.nih.gov
odds ratio	The ratio of the odds of an event occurring in one group to the odds of it occurring in another group, or to a sample-based estimate of that ratio. An odds ratio of 1 indicates that the condition or event under study is equally likely in both groups. An odds ratio greater than 1 indicates that the condition or event is more likely in the first group. And an odds ratio less than 1 indicates that the condition or event is less likely in the first group. The odds ratio must be greater than or equal to zero. As the odds of the first group approaches zero, the odds ratio approaches zero. As the odds of the second group approaches zero, the odds ratio approaches positive infinity.
p-value	In the case of SNP disease association studies, a statistical measure of evidence that the SNP is associated with the disease phenotype
SNP	Single nucleotide polymorphism: A SNP occurs when corresponding sequences of DNA from different individuals differ at one DNA base; for example, where the sequence AAGCCTA changes to AAGCTTA.
SNP panel	A collection of SNP loci genotyped together on a genotyping platform (e.g. Illumina)
whole genome rank	The rank of significance (disease phenotype association) of the specific SNP in the analysis (the lower the rank, the higher the significance)

INDEX

U

W