



Searching for SNP Genotype, Allele and Frequency Data

Created: July 7, 2005; Updated: February 25, 2014.

Searching for SNP Genotype Data

Is there a way to get a genotype data dump file for selected SNPs, or SNPs in a certain region?

You can use dbSNP's [batch query](#) service. (03/08:02/14)

I am unable to find SNP data from the second phase of the HapMap Project recently published by the HapMap consortium.

The [phase II HapMap genotype data](#) (see the [PubMed](#) abstract of the phase II data publication in Nature) is currently available in dbSNP.

This genotype data is from HapMap release 21a which contains both phase I and phase II data. There are a number of ways to access this data from dbSNP: you can use [Entrez SNP](#), [eUtils](#), or an [FTP](#) download depending on whether you need the entire data set or data specific to particular genes or chromosome regions. (11/27/07)

After narrowing the range of my search for SNPs that are polymorphic between the C57BL6/J and 129 sv/j strains of mice by typing 3[CHR]105351425:107237468 [CHRPOS] into EntrezSNP, How do I get Entrez SNP to show me SNPs that are polymorphic between C57BL6/J and 129 sv/j?

Enter the following into Entrez SNP:

```
"Mouse[orgn] AND true[gtype] AND 3[CHR] AND 105351425:107237468 [CHRPOS]"
```

(11//06:02/14)

How do I upload the data for only those SNPs that have genotype or frequency information?

There are two ways to upload the data for all SNPs that have genotype or frequency information.

The easiest way to do this is to upload the Genotype XML GenoExchange files. These files contain all dbSNP Genotype and allele frequency information arranged by chromosome for each organism. For example, the [human build 125 genotype data](#) are located in the genotype subdirectory of the human SNP directory. The schema for the GenoExchange files is also located [online](#).

Another way to upload the data for all SNPs that have genotype or frequency information that requires a little more work, is to upload all the dbSNP database files, which are also available on our [FTP site](#). (5/6/06)

How do I find the genotypes for rs2074192 in dbSNP?

To view the genotypes of a single SNP:

1. Go to the [SNP home page](#) and find the "Search by ID" section. It is located below the announcements.

2. Type "rs2074192" (don't include the quotes) into the Search by ID text box, then click on the "Search" box. This will generate a refSNP cluster report for rs2074192.
3. Now, scroll down to the Variation Summary" section of the refSNP cluster report and click on the words "genotype detail" link located in the left hand side of the section. This will display the genotype report for this rs number. (1/25/05)

How do I download genotype and functional data in the XML format?

Genotype and functional data are stored as modules in two separate XML reports, so you'll have to download using eUtils and then combine the results from the two reports.

Below is an example of how to use eUtils to retrieve human SNPs that contain genotypes in the LPL gene:

1. [eSearch](#) using the term "LPL AND true[gtype] AND txid9606".
2. Parse the XML results to get a list of refSNP (rs) ID numbers.
3. [eFetch](#) to retrieve the reports for each SNP. (2/7/06)

I plan on using a Coriell panel of DNA samples from 50 individuals as controls in a genotyping study. Can I search dbSNP for single as well as multiple SNPs located in a single individual's DNA?

The dbSNP genotype report can be used to determine if a given SNP (or list of SNPs) has submitted genotypes for the samples you have purchased. Using the batch query option may be the best way for you to do this:

1. Go to the [dbSNP home page](#) and scroll down until you find the "Search" section heading on the blue left side bar, then click on the words "Batch Query", located four lines beneath it.
2. Scroll down to the bottom of the Batch Query page, and click on the "submission format" drop-down menu toggle to see a list of Batch Query submission options. Select "Enter RS#"
3. Now, enter your email address in the text box in the "Email" section of the page, select the appropriate organism from the drop-down menu in the "Organism" section, and enter your refSNP numbers along with their rs prefixes in the text area of the "Enter RS Numbers" section.
4. From the drop-down menu located in the "Select Result Format" section, select "genotype report". Click the "Submit" button.

An xml file containing all submitted genotypes for each of the refSNP numbers you entered will be emailed to you. The report will contain all populations and individuals typed for the SNPs you entered into the batch query since there is no option to select only specific individuals or populations at this time.

You can look at a [schema document](#) that contains details on the xml format for the genotype report.

You can also look at the SNPs one at a time by doing the following:

1. Go to the [dbSNP home page](#) and enter a rs# in the textbox at the top of the page, and click the "Go" button.
2. Click on the red "G" link shown in the resulting graphic to bring up the genotype report in html.
3. Click on the "+" sign next to a listed population name to expand a table of genotypes. In the tables, each row represents a sample. (6/16/05)

How can I get a dump of the new AFFY genotypes?

Go to the [dbSNP homepage](#) and click on By Submitter in the blue sidebar (under Search). Type "AFFY" in the textbox and then click the Search button. Affymetrix submissions will be displayed. Now, click on the submitter's handle to view the contact detail and obtain a list of the batches. Click on batch_id to see the populations. Finally, click on detail under view genotype to display the actual genotypes.

The Genotype Query Form

How do I query and download genotype data now that the Genotype Query Form has been retired?

There are three alternative sites you can use to query and download genotype data:

- 1000 Genomes: <http://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/>
- HapMap: <http://hapmap.ncbi.nlm.nih.gov/>
- Genotype XML on the dbSNP FTP site: <ftp://ftp.ncbi.nih.gov/snp/organisms/>

The genotype XML provides both submitter and genotype information for many submitted SNPs, and is organized in chromosome specific files under each organism directory in the “genotype subdirectories”. For example, human genotype XML files are located in ftp://ftp.ncbi.nih.gov/snp/organisms/human_9606/genotype/.

Why was the Genotype Query Form retired?

Because the large volume of genotype data we receive from large sequencing projects (e.g. 1000 Genomes) makes it difficult for NCBI to maintain and query the dbSNP SQL tables, NCBI is currently developing an alternative genotype database using another technology that will more efficiently store and serve genotype and frequency data. It should be available sometime in 2014.

Will dbSNP release a new Genotype Query Form?

We are currently developing an alternative genotype database using another technology that will more efficiently store and serve genotype and frequency data. It should be available sometime in 2014.

Searching for SNP Allele Data

Is there a way to determine if a triallelic SNP in dbSNP was found in a single individual?

When a triallelic SNP is submitted to dbSNP, we do not know if they were ascertained in a single individual or a pooled population. For some of the triallelic SNPs, there maybe associated individual genotype or allele frequencies that might determine whether it was in a pooled population.

During [the SNP merging process](#), co-located SNPs with different biallelic pairs (e.g. A/T and G/T) are collapsed into a single triallelic observed line (e.g. A/T/G). (05/13/08)

Can you provide a search strategy for finding tri-allelic and/or tetra-allelic SNPs in dbSNP?

dbSNP does not have a simple web search for your question. You'll need to download the database table dumps located on the [ftp site](#) and use SQL to extract the data. The dbSNP [database ER Diagram](#) will help you interpret the tables.

The query is complex so let us know if you need help. We'll look into providing such a search in Entrez SNP. (11/30/07)

Searching for SNP Frequency Data

Finding Expected Frequency of Multiple SNPs

What is the expected frequency of eight SNPs occurring in the coding region of a single, 1,300 bp ORF?

The frequency of multiple SNPs is estimated to be two exonic SNPs per gene (coding and untranslated regions). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms is located [online](#).

Suggestions for verifying your variations:

Sequence cDNA products from several independent PCR reactions or test multiple clones.

Use a different cell line for the control reaction.

Finding Allele Frequency Data

How do I access allele frequencies in dbSNP?

There are several ways to get allele frequency information:

- Use the [average frequency information](#) for a single reference SNP (refSNP, rs#).
- The [frequency information for a single submitted SNP](#) (ss#), is located toward the bottom of the page. The same data is also available in [submission format](#).

To view genotype information for a particular population, do the following:

1. Go to the dbSNP homepage and select the Search by Population detail link and conduct a search as indicated in the documentation.
2. Once dbSNP generates a report of the population detail, select a sub-batch ID of interest to generate a Population Detail report.
3. To generate an Individual Genotype Batch report, select a batch ID of interest from the Population Detail report.
4. To view the individual genotypes of your batch, Select detail located under View genotype in the batch summary section. An example of the [Individual Genotype Batch report](#) is available.

Allele frequency data are available at dbSNP's FTP site in several formats—XML, asn.1, and submission format.

You can also obtain the average allele frequency for all rs numbers in the [SNPAlleleFreq](#) table located in the organism_data directory for your organism.

Finding Minor Allele Frequencies

Is there an automated way to retrieve minor allele frequency for all mouse SNPs?

As there are over 14 million mouse SNPs, this would be too much data for a web-based query and retrieval, so you'll need to download the [Allele](#) and [SNPAlleleFreq](#) table from the ftp site and query it locally.

Descriptions of the tables can be found using the [dbSNP Data Dictionary \(08/27/08\)](#)

Could you tell me the minor allele frequency for rs1050622?

This SNP is monomorphic in a CEPH population, according to the frequency data that were submitted for this SNP.

You can see this by looking at the "[Population Diversity](#)" section of rs1050622's cluster report on the rs report. [\(06/26/08\)](#)

Finding Allele Frequencies across Major Populations

Is it possible to download just the SNP frequencies for all SNPs across the European Caucasian population in a table format?

We currently do not have a flat file report of allele frequencies across major populations for all human SNPs. If you are able to work with XML files, there are FTP files of genotype data ([/organisms/human_9606/genotype/](#) and [/organisms/human_9606/genotype_by_gene/](#)), which are organized by chromosome that include genotype and allele frequency data. [\(11/28/08\)](#)

I want to download flat files that list all SNPs for which allele frequencies exist across the major human populations (European, East Asian, and African). Where can I find such data?

We currently do not have a flat file report of allele frequencies across major populations for all human SNPs. If you can work with XML files, you can find FTP files of genotype data (including genotype and allele frequency data) in the dbSNP FTP site. They are organized by chromosomes:

ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/genotype/

ftp://ftp.ncbi.nlm.nih.gov/snp/organisms/human_9606/genotype_by_gene/

If you are working with a set of SNPs, you could use [eutils](#) to fetch the allele frequency information.

In the meantime, we will look at making a new FTP file format for reporting SNP frequencies. (7/24/07)

Finding Population Frequency Data

How do I find SNP allele frequency information for different populations?

You can get SNP allele frequency for different populations by using the genotype report format. To get this format, select a refSNP from a population you are interested in, let's use rs221 for this example.

1. Go to the Entrez SNP page and type in rs221 in the search text box at the top of the page, and click the "Go" box to generate the rs221 record.
2. Click the "Display" drop-down menu toggle located just underneath the tabs on the left near the top of the SNP page, to see a list of SNP display options. Select "genotype detail" as the display type
3. If you click on the "+" sign located just to the left of each genotype option, you will see SNP frequency for populations available on dbSNP. You can also get this data in xml format by choosing "genotype XML" as the display type in step 2. (5/4/05)

How do I find population data on SNP frequencies posted on dbSNP? I am looking for large, random, ethnically defined population frequency data and do not have a PMID number.

Because you mentioned "PMID", I'm assuming that you have a publication title or an author's name. If so, you can [search dbSNP](#) using the publication title or the author's name.

Clicking on the title leads you to the publication page. At the bottom, there is a list of all batches citing this publication. [Here's an example](#). Look for batch_type "frequency".

The above approach will not give you all SNPs with large, random, ethnically defined population frequency data, but if you have a publication in mind, it will give you a start.

How do I identify all dbSNP frequency data coming from the HapMap CEU population?

The HapMap Handle is "CSHL-HAPMAP" and the HapMap CEU population ID is "HapMap-CEU". Using the dbSNP "[Population Detail](#)" search, enter HapMap-CEU in the test box in the grey query section, and then select "submitter population ID" and "exact". Click on the "HapMap-CEU" link you get in your response to get the [details](#) for the HapMap-CEU population.

One way to get the allele frequency information for this population is to parse the genotype and allele frequency (genoExchange format) xml files found in the [human genotype directory](#) of the dbSNP FTP site. You can find documentation for the genoExchange format [online](#).

All of the "ByPop" elements that have the attribute pop_id="1409" are from the HapMap CEU population. (11/06:02/14)

Finding Populations used to Determine Frequencies

How do I find the populations used to determine SNP frequencies, and whether these frequencies vary by ethnic group?

Follow the example below:

1. Go to [Entrez SNP](#)
2. Type in "LPL" in the input text box at the top of the page, and click the "GO" button located to the right of the input text box.
3. You will get a report that has two sets of tabs at the top. Look at the lower set of tabs and click on "Human" tab.
4. You will see a list of rs ID numbers. Scroll down the list of rs ID numbers and look for one that has a red "G" located on the right end of the graph bar.
5. Click on the red "G" to show genotype details for that rs ID number. If the rs ID does not have a red "G", then no genotype detail is available for that rs ID

Another way of finding genotype information for an rs number is to use the "Genotype Detail" link located in the "Variation Section" of any refSNP report:

Follow the example below, using any refSNP page of interest.

1. Scroll down to the bottom of the report, where you will find the variation section.
2. Click on the blue "Genotype Detail" link at the bottom of the variation section
3. At the bottom of the list of genotypes, you will see "SNP Detail" bar. Find and click on the blue "+" link located at the front of this bar.
4. You will see frequency detail organized by each submitted SNP(within the refSNP cluster)and population.
5. If you select the "xml" format from the "Display" drop down list at the top of the page, you will also get the computed Hardy-Weinberg Probability (10/04/05)

Uploading Data for only those SNPs with Genotype or Frequency Data

How do I upload the data for only those SNPs that have genotype or frequency information?

There are two ways to upload the data for all SNPs that have genotype or frequency information.

The easiest way to do this is to upload the Genotype XML GenoExchange files. These files contain all dbSNP Genotype and allele frequency information arranged by chromosome for each organism. For example, the [human build 125 genotype data](#) are located in the genotype subdirectory of the human SNP directory. The schema for the GenoExchange files is also located [online](#).

Another way to upload the data for all SNPs that have genotype or frequency information that requires a little more work, is to upload all the dbSNP database files, which are also available on our [FTP site](#). (5/6/06)