U.S. National Library of Medicine
National Center for Biotechnology Information

SNP FAQ Archive
NCBI Help Manual

National Center for Biotechnology Information
U.S. National Library of Medicine

Choosing the Data Report that Contains the Data You Need

Created: July 7, 2005; Updated: February 18, 2014.

Which dbSNP report will give me a list of refSNP numbers, their corresponding chromosome number, base pair position, strand and alleles?

The closest format we have that would provide you with the information you are looking for is the Flat File (FLAT).

If this is not suitable, you'll have to extract the data from an XML file. (**06/11/09**)

I work with many thousands of SNPs, and would like to visualize information on a large scale. Do you have a format that will hold a lot of data for each SNP in a compact fashion?

SNP has its data available for many organisms as a bitfield, which can compactly hold a great deal of information. For humans, you can access SNP_bitfield.bcp.gz in the human subdirectory of the organism_data directory. The specs for SNP_bitfield located in the specs directory of the dbSNP FTP site.

The bit field can be retrieved using the NCBI C++ Toolkit with its feature iterator and annotation selection classes. This would imply that you have developed a C++ program that will read the bit field too. The bit field can also be retrieved from raw XML or ASN.1 dumps taken from Entrez. You can use the PERL programming language to extract the bit field from there without using C++. (**11/08/07**)

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**)

Where can I find allele frequency and population information for submitted SNPs? I can't find them in the B125 xml file for my refSNP.

Use the genotype report for population and allele frequency data. (**11/17/05**)

Does dbSNP have a single flat file available for download that contains the chromosome assignment(s) and physical positions(s) for each rs entry?

Yes. There is a flat file called the "Chromosome Report", located in each the dbSNP FTP site for each organisim. "Chromosome Report" provides an ordered list of RefSNPs in approximate chromosome coordinates (the same coordinate system used for the NCBI genome MapViewer). The column definitions for

"Chromosome Report" are located in the dbSNP FTP readme file. Once in the readme file, scroll down until you come to the "Chromosome Reports" section. **(12/01/05)**

Which dbSNP report format contains the location and other important information for each SNP?

The "chromosome reports" (chr-rpts) report format located in the dbSNP FTP site reports the map locations of all SNPs in both contig and chromosome coordinates. You can access this file using the following steps:

1. Go to the dbSNP FTP site, and click on your organism of interest. You will be taken to an index of the reports available for that organism. Please note that not every organism has data available in the "chromosome reports" report format.
2. Click on chr_rpts. This will take you to a list of the available chromosome reports, listed by chromosome, for the organism of interest. Please note that not every chromosome for a particular organism may have a chromosome report.
3. Documentation for the columns found in the "chromosome reports" report format can be found in the general read-me file, under the heading "CHROMOSOME REPORTS".

Is there a gzipped flat file in dbSNP's FTP that contains refSNP numbers, locations (bp), chromosomes, HUGO, functional classes, and amino acid changes for SNPs on multiple chromosomes?

Go to dbSNP's FTP site, and select the "organisms" directory. Once in the Organisms directory, select the organism you are interested in (I'll choose human for this example). Once in the appropriate organism directory, select the "ASN1_flat" file. (**8/8/07**)

Where can I find a flat file that contains all human SNPs, including those SNPs that are not located in genes?

The ASN1_flatfile is only flat file source that contains all human SNPs, (organized by chromosome) including those SNPs not located in genes.

For those SNPs not located in a gene, you can conduct this search on Entrez SNP using the following search terms:

"(txid9606[All Fields] NOT snp_gene[sb]" (do not include the quotation marks).

Or, you can download all the files from the ASN.1 flat file and parse out the SNPs not located in genes.

See the eUtils documentation if you want to retrieve the SNPs programatically. (**08/09/07**)