



Eutils (E-utilities)

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

Using eutils to Obtain Specific SNP Information

Can I use eutils to get SNP location type, class, validation, function, molecular type, heterozygosity and weight for all the SNPs on a chromosome?

As it would take entirely too long to get this information for 14 Million+ SNPs by eutils, you should get the information from the dbSNP FTP site. If you need to look at all the SNPs on a particular human chromosome, go to the [organisms/human_9606/XML/](#) directory on the dbSNP FTP site, and select the chromosome file you would like to see. (10/06/08)

How do I use eutils to extract from dbSNP some fields that are usually in the SNP annotation of the Illumina repositories? These fields include CytogeneticBand, GeneticMapPosition, LocationRelativeToGene, etc.

Sorry, dbSNP doesn't have the Illumina fields you mention available in eUtils. The SNP data available in eUtils are documented in the [docsum schema](#). (10/25/07)

How do I extract information from dbSNP by using eutils and specifying the chromosome position interval?

Use the following eSearch query: "100000:200000[Base Position] AND "1"[CHR] AND txid9606". See [example](#).

You can also get the result using [eFetch](#). (7/9/07)

Eutils Workshops and Tutorials

Where can I get more information on NCBI workshops for data retrieval?

There is [online information](#) for courses offered by NCBI.

You might also want to go to the [Eutils General Help page](#), and look at the NCBI online short course book called "Building Customized Data Pipelines Using the Entrez Programming Utilities (eUtils)" (12/01/05)

Is there an online tutorial for eutils?

There is an [online tutorial](#) for using eutils on the NCBI Bookshelf (9/20/05)

Eutils Online Documentation

Is there some kind of help documentation for eutils?

There is an [online book](#) that describes how to use eutils (2005)

Using Eutils to Find Specific Data

Using Eutils to Find rs Merge Information

Can I use eutils to get the information about a particular rs number to include a notice that the rs number has merged into another rs number?

Currently, there are three different entry points in dbSNP that will lead you to the partner numbers of a merge:

1. You can retrieve a list of merged rs numbers from [Entrez SNP](#). Just type “mergedrs” (without the quotation marks) in the text box at the top of the page and click the “go” button. Each entry in the returned list will include the old rs numbers that has merged, and the new rs number it has merged into (with a link to the refSNP page for the new rs number). You can limit the output to merged rs numbers within a certain species by clicking on the “Limits” tab and then selecting the organism you wish from the organism selection box.
2. If you enter an old rs number (in this case rs2857713) into the “Search for IDs” search on the dbSNP home page, the response page will state that the SNP has been merged, and will provide the new rs number (in this case rs2229094) and a link to the refSNP page for that new rs number.
3. The [RsMergeArch](#) table houses the merged SNPs, and is available on the dbSNP ftp site. A full description of the table can be found in the [dbSNP Data Dictionary](#), and the column definitions are located in the dbSNP_main_table.sql.gz, which can be found in the [shared_schema](#) directory of the dbSNP FTP site.

(11/03/08)

Using Eutils to Retrieve Population Diversity Files

Is it possible to use eSearch and eFetch to get population diversity XML files?

You can use eSearch to search for SNPs that have genotype information (please see this [example](#)), and eFetch to retrieve the genotype report to get the population information. You can find additional report examples on the “[Eutils for EntrezSNP](#)” page.

(11/17/08)

eFetch

How do I use efetch with EntrezSNP?

There is online [documentation](#) for using efetch with EntrezSNP. (5/13/05)

Can I use ePost and eSummary or eFetch to get SNP records in various formats?

You can retrieve SNP records using eFetch in various report formats if you already know the snp_id. (e.g. 3, 7775). Please take a look at our [eutils page](#) for more information.

Here is an [example](#) of one of the report types delivered using efetch and the ID number mentioned above.

(4/30/07)

How do I fetch data for a given refSNP ID (rsID) in XML format using a simple URL-request?

Please see the [online short course](#) for using eutils (9/15/06)

Is it possible to use eSearch and eFetch to get population diversity XML files?

You can use eSearch to search for SNPs that have genotype information (please see this [example](#)), and eFetch to retrieve the genotype report to get the population information. You can find additional report examples on the “[Eutils for EntrezSNP](#)” page.

(11/17/08)

eSearch

Will the HGVS based variation search you are testing also be available through eutils?

Yes, it should also be available through eutils once we have finished developing the resource (9/27/07)

If I query eSearch for all SNPs that hit on chromosome 1 in the 1 to 6000 bp range, I get 11 hits, but if I search our local MSSQL directly, I get 115 hits. Why the difference? How do I get the same results that I got from Entrez by using a direct SQL query, and vice versa?

The answer to your first question is that Entrez only returns SNPs that clearly map once to the current genome build, while a direct SQL query will deliver SNPs that map more than once.

To get the same results that are returned by Entrez by using a direct SQL query, that is, to get an refSNP (rs) number with "clear", or as we call it, a "unique" map hit, please see the Database Dictionary description for [SNPContigLoc](#), and for [SNPMAPIInfo](#) (look for the weight information in SNPMAPIInfo).

To get the same results that were delivered by SQL using an Entrez query is not really possible. When a SNP maps to multiple locations on multiple chromosomes, Entrez indexes the SNP as chrMulti. So if I use your example of finding SNPs with hits on chromosome1 in the 1:6000 bp range, you will find in SNPContigLoc all SNPs (including those with weight higher than 1) that hit in the 1:6000 range.

But if you use Entrez, however, you will find only weight 1 SNPs (weight 1 represents SNPs that align at exactly one locus) that hit chromosome 1 in the 1:6000 range. SNPs that are weight 2 (weight 2 represents SNPs that align at two loci on same chromosome) are not indexed in Entrez, so you will not find the same number of SNPs that hit chromosome 1 in the 1:6000 range using Entrez as you do using a direct SQL query.

(8/27/07)

XML Format for eutils

Why does dbSNP have two different XML formats (one for the file and one for data for Eutils) for SNP entries?

We have plans to merge the two XML schemas in the near future. As for the schema merge it will be done some time this year. As we have a lot of ongoing projects currently, I'm unable to be specific about when this will be accomplished. (2/5/07)

I'm using the command “<http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch>

.fcgi?db=snp&rettype=genxml&id=\$marker” to extract freq info from dbSNP, but can't get well formed XML.

Please add “retmode=text” (no quotes) to the command: <http://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=snp&rettype=genxml&id=3000&retmode=text> (09/05/08)