



Batch Query: Retrieving information for a large numbers of SNPs

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

How do I BLAST dbSNP? It used to be easy. Yikes, I can't figure it out!

Go to the [dbSNP Home page](#), and select "search" from the left blue sidebar. A dropdown menu will appear. Select "Blast SNP" to go to the [BLAST page](#).(2/20/07)

Conducting BATCH Queries using Specific ID Types

How do I submit a list of SNPs in the following format #[CHR]StartPos:EndPos[CHRPOS] ex- 2[CHR]123456:765432[CHRPOS] and get all the SNPs present in the specified coordinates on the chromosome number?

You'll have to retrieve the data programmatically using [eUtils](#). You can use [eSearch](#) and [eFetch](#) for searching for and retrieving reports. An [overview](#) of all Entrez Programming Utilities is available online. (6/5/07)

Can I conduct a Batch query using Celera ID numbers?

The dbSNP batch query service now accepts hCV numbers. You can either enter the numbers in the text box entry site on the web, or you can upload a list from your computer.

1. Go to the [dbSNP Batch Query](#) site.
2. Scroll down to the text box with "Submission Format" in it, and click on the blue arrow to activate a dropdown menu.
3. Select ""Upload Celera ID" to upload a list from your computer, or select"Enter Celera ID" to enter the Celera ID numbers directly.
4. Follow the directions for entry given on the response page.

When you are ready to enter or upload your list, make sure that the list is formatted with one ID per line as shown below:

HCV1

HCV100

HCV100000

HCV1000000

HCV1000004

HCV100001

HCV1000012

HCV1000013

(4/10/06)

When I run a Batch query using Celera/ABI SNP IDs, the Chromosome report I get back doesn't contain the original query IDs, and is sorted by rs number, so I've lost correlation between the IDs. What can I do?

Only the FLATFILE report lists the submitter's ID (i.e. in your case, the hCV IDs) of all submitted SNPs. hCV IDs, however, are not reported for SNPs that are submitted by an entity other than Celera.

dbSNP does not have control over the assignment of specific rs numbers to hCV IDs. Not every rs number has a corresponding hCV ID, and Celera may change or drop the rs number based on their internal processing. Likewise, dbSNP's mapping and clustering process may change, merge, or drop rs numbers. I suggest that you convert your processing from hCV to dbSNP rs numbers, since rs numbers are the standard identifier for SNP.

There is a table of hCV IDs and their corresponding rs and ss numbers (for dbSNP build 127) located in dbSNP's FTP site in the [misc. folder](#) of the human_9606 directory. The file name is hCV_dbSNP_ID.txt.gz. The column definitions are located in the hCV_dbSNP_ID readme file located in the [same directory](#) and folder. (2/13/07)

Retrieving Specific Data using BATCH Query

Retrieving Chromosome data

I need to download all the SNP chromosome positions in build 125 and 126, but the chr_rpts files I downloaded were the same for each build I selected.

I assume you want human build 125 and 126 map positions, right? You could get them in the [organism_data subdirectory](#) for each organism (the link above goes to the human organism_data subdirectory). **Please Note:** the files you are looking for start with the build number followed by "SNPContigLoc"; e.g. b125_SNPContigLoc_35_1. The column definitions for SNPContigLoc are located in dbSNP's [Database Dictionary](#). (03/07/08)

How do I get the chromosome, chromosome position, alleles and Ancestral allele for each SNP in a file of refSNP numbers I obtained from HapMap?

You can use dbSNP [Batch Query](#) service to download the Flat File Report format for a list of rs numbers. The Flat File Report contains chromosome, chromosome position, and allele information for each refSNP. You may also wish to look at the [ancestral allele section](#) of the SNP FAQ archive for further information on ancestral alleles. (9/19/06)

When conducting a BATCH query, how do I specify a specific set of fields for the query to retrieve (chromosome positions, rs#, and gene name)?

On the Batch query submission page select "Chromosome Report" as the result format. (4/10/06)

How do I use dbSNP refSNP (rs) numbers to get a list of chromosome positions?

1. Go to the [dbSNP home page](#), and scroll down to the "Batch" section.
2. In the "upload list" sub-section, select "Reference SNP ID (rs)".
3. On the data request form, enter the email address where you want your data sent, select the organism of interest from the "organism" dropdown menu, and in the "Select Result Format" section choose "CHROMOSOME RPT" from the dropdown menu (3/1/06).

Retrieving Functional Data

Is there a way I can download the functional information for a list of about 60,000 SNPs in text file format?

You can upload your list of refSNP numbers (30,000/per load) to the [batch query](#) service and request the “FLATFILE” report.

Click this [link](#) to see the FLATFILE report for rs1855025. (5/12/05)

Retrieving Frequency Data

How do I get SNP frequency information using a batch query?

Frequency information is located in the genotype report. Please use the batch query and select genotypeReport.

Converting Local IDs to rs and ss IDs

How do I use dbSNP batch query to convert a series of SNPs from a local lab to rs and ss numbers?

Create a list using only the local SNP ID, but don't include the handle.

Retrieving FASTA Data

I can't get Batch Query to retrieve FASTA reports for the submitted SNPs (ss) I submitted several months ago.

Batch Query currently only works for those SNPs that have been clustered into a RefSNP (rs), even though you are only asking for a submitted SNP (ss) report. We understand that this is a problem when there is a substantial lag time between the ss submission and rs clustering. In December of 2008 we instituted a new Batch Query feature that allows a Batch Query search for pre-clustering ss fasta reports. (10/09/08)

Can you send me the FASTA sequence for a list of rs numbers?

You can use dbSNP [Batch Query Service](#) to download the FASTA format for a list of rs numbers. (9/19/06)

Retrieving SNPs, their frequencies, function, chromosome positions and FASTA files

How do I retrieve all the reported SNPs, their frequencies, function, chromosome positions and FASTA files for 50 genes? Would I use a Batch search? If so, how?

The Batch query is not the correct way to proceed. Instead, you will have to write a program using the [eutils programming utilities](#):

1. Use eSearch to perform the search, and then parse out the refSNP (rs) numbers you need. Here are the [results](#) for a search for SNPs in the human LPL gene.
2. Use eFetch to retrieve the SNP reports (6/20/06)

Retrieving SNPs in Specific Genes

How do I use BATCH query to obtain detailed information in FLATFILE format for all of the SNPs in 200 genes?

dbSNP doesn't have a BATCH query for Gene ID. You'll have to use eSearch and eFetch to do it programmatically. There is another [FAQ](#) in this archive that shows you how use eSearch and eFetch in this way. (5/22/08)

Using BATCH to Retrieve Unclustered Submitted SNPs (ss)

How can I get information for a large number of submitted SNPs (ss) whose refSNP (rs) numbers have yet to be assigned, and will therefore not generate XML or FLATFILE reports in response to a Batch query?

Use the following instructions to upload the ss numbers of interest using the Batch query service and request the Submitted SNP “SubSNP” details report format:

1. Go to [dbSNP Batch Query](#) page.
2. Scroll down to the submission format drop-down menu and select “Upload SS#”.
3. Enter your email address in the appropriate text box, and select the organism of interest from the organism drop-down menu.
4. Select "SubSNP Details" from the Result Format drop-down menu.
5. Upload the series of ss numbers of interest using the “Browse” button. Once you have uploaded the ss numbers, Click on the “Submit” button.

The Batch query limit is 30,000 SS numbers per request. (4/4/06)

BATCH Reports

BATCH Reports compatible with Excel

Is it possible to convert a BATCH report into an Excel spreadsheet? I want to use the Excel search function to query using a refSNP (rs) number so I can easily find the gene name, variation, function, and frequency information for the refSNP.

When you are entering your BATCH query, select the chromosome report as your result format. The chromosome report format is a tab-delimited table that can be imported into Excel. You’ll have to use `efetch` to get the allele frequency from the XML files (FREQXML), however. Remember that NOT all SNPs have frequency data. Examples of the different dbSNP formats retrieved using `efetch` located [online](#). (7/28/06)

BATCH Queries and XML Formatted Reports

How do I extract data from XML output for rs batch queries?

I would recommend that you use an XML parser because of the complex structure of XML. There are parsers available for most common computer languages. You can see an [example of parsers used in PERL](#). Another option is to [use XSLT](#) to extract and transform XML to the format of your choice.

When using XML output for rs batch queries, how do I select the frequency data for specific rs numbers from among the frequency data listed for all the ss numbers?

To get the rs allele frequency directly, use the FTP file, `SNPAlleleFreq.bcp.gz`, located in your organism’s `organism_data` directory (ex: [human](#)).

The fields in the file are tab delimited and are defined in the schema table for your organism (ex: [human](#)).

If I need to get more than 40 rs numbers from XML-formatted genotype and allele frequency reports, will I need to write a Perl script to do so?

Here are instructions for obtaining an genotype exchange XML file for a set of specific SNPs:

If you have a list of SNPs (rs numbers), you can upload the list to batch query by using the [rs page](#). At the rs page, type in your email address. Then select `genotypeReport` as the format and click on submit. The report will be emailed to you.

Alternatively, if you want a large list of rs numbers and want to start from a Entrez query, do the following: Once you have narrowed your Entrez query to the list that you want, select the dbSNP Batch report option, click display, and follow the instructions in the rs list example.

In addition to the SNP-specific queries, we are currently exploring various options for allowing users to specify population, pedigree, and/or individual filters as limits to the downloads for the genotype exchange XML report.

Sequence Data provided in BATCH Reports

Will the results for a BATCH query that I've submitted contain SNP flanking sequences from various builds or will the flanking sequences be from the current build (b125)?

The flanking sequences will be from the current build — dbSNP build 125, which is based on human genome build 36.(5/10/06)

BATCH Query Limits

I need data that shows the conversion between Baylor IDs and rs IDs, but the Batch Query limit is 30,000 SNP IDs, which means I'd have to run 56 batch queries. Is there an easier way?

You can download the [SubSNP](#) table from the FTP site and use that to lookup Baylor IDs using dbSNP rs IDs.

The [column definitions](#) for the SubSNP table can be found using dbSNP's [data dictionary](#) (07/08/08)

I would like to get Chromosome Reports for a whole genome SNP panel (100K). How many SNPs can BATCH query process at a time?

Each BATCH Query is limited to 30K SNPs. (12/08/07)

Is there a maximum number of SNPs that can be uploaded to a Batch search?

The dbSNP Batch Query has a limit of 30K of refSNP (rs) or submitted SNP (ss) numbers per batch. (3/5/07)

BATCH Query Errors

Unable to use Entrez BATCH upload

I am trying to retrieve allele frequencies for 2500 SNPs, and would like to get a text file with the results directly, but I get an error when I try a BATCH upload.

Try using the [non-Entrez Batch query](#) service and requesting the genotype xml report which include allele frequency. We'll look into the generating a simpler frequency report. (03/26/08)

Unable to Access Old Batch Query Results

I have a batch query I want to access but cannot gain entry to the FTP server to retrieve it.

The file you're trying to access was created in 2003 and has been deleted from the system. We cannot keep files in the FTP site for than 48 hours because of limited resources. Please submit your query again to generate a new file.

Unable to Access Recently Submitted SNPs using BATCH

I've tried to batch download pig SNPs recently added to dbSNP but couldn't get anything besides a list of the SNPs to appear on the dbSNP webpage, and the "Batch Query Result" email attachment I was sent was empty. Why?

The batch query service only dumps reports for SNPs that have been clustered and assigned with reference SNP (rs) numbers. The pig SNPs you are inquiring about have yet to be clustered and assigned with reference SNP (rs) numbers.

I downloaded 30 submitted SNP (ss) accessions from the SubSNPAcc.bcp.gz file and submitted them as a batch query, but I have yet to receive a fasta file back.

The batch query does not work for unclustered SNPs, and since the records you submitted represent submitted SNPs (ss) that were not mapped and/or clustered, you did not receive a response. You can download the records for these SNPs from the [dbSNP FTP site](#).

(7/18/06)

Unable to Access XML/FLATFILE Formats using BATCH

When I enter a series of submitted SNP (ss) numbers into a Batch query, select *Homo Sapiens* selected as the organism, and select FLATFILE or XML as the report format, the query generates no information.

Only submitted SNPs that have had refSNP (rs) numbers assigned to them can be as can be released in the FLATFILE or XML report formats. The submitted SNP (ss) numbers you submitted for a Batch query were only just recently assigned to some new submissions that came into dbSNP, and therefore have not been assigned refSNP (rs) numbers yet. The rs numbers will be assigned by the next dbSNP release. (3/29/06)

Local ID Upload Error

During upload of a list of Local SNP IDs, I get an error message stating the EMAIL, REPORT_FORMAT and QUERY_FORMAT fields were missing, but they're not.

Make sure that the file you upload contains one SNP per line with the submitter handle and Local SNP ID delimited with a pipe "|" (e.g. TSC-CSHL|TSC0000026).

Also, please note that the maximum upload for each file is 30K IDs or lines. The query will fail if you exceed this limit.(10/06/08)

Local Email server Deletes Emails with zip/.gz Attachments

Our email server strips away emails that contain attachments ending in .zip or.gz. Is there another way to retrieve batch reports?

You can use dbSNP's Entrez batch query system if you're querying with refSNP (rs) numbers. The Entrez SNP batch query system is fast, but it can accept only rs numbers, and as such you need only submit the number without the "rs" prefix attached.

1. Go to the [SNP batch Entrez page](#)
2. Click on the "Browse" button at the top of the page, and select the rs list you intend to query.
3. Click on the "Retrieve" button located next to the "Browse" button
4. When the result is shown, select the "Display" option and select a report from the list of available report types.
5. Select the "Send to" option to save or print the report.(5/19/06)