



Retrieving Specific Data for a SNP

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

Finding Contact Information for an Individual Submitter

Using the links provided in the submission format report, I found the handle, and therefore the contact information, of the lab that submitted the SNP. The problem is that I need the contact information for the actual individual in the lab who submitted the SNP.

As you stated in your question, each Handle is associated with the contact information for the Principle Investigator, lab chief, a project leader, or group (lab). When a person from the lab (or group) submits, they are supposed to put the lab's Handle in the Handle field of the submission, and insert their own name, telephone, fax, and email information in the telephone, fax, and email fields located below the Handle field. The problem is, however, that much of the information that goes into the name, telephone, fax and email fields is often the same as that of the group/lab the submitter works in, so the submitting individual will forget to update the name, telephone, fax and email fields with their own contact information, and as a result, when you search for the name and contact information of the individual within the lab (or group) who submitted the SNP, you get the information for the lab. Our submission templates were not clear on this point, and have been updated. (7/18/07)

Finding out which SNPs are Repeated and which have Multiple Submissions

How do I use Entrez SNP to determine which SNPs are repeated and which are multiple submissions?

Click on the "Limits" tab and select "by-submitter" under Validation Status. (06/23/08)

Finding SNP Citations (SNP Publication Details)

How can I find the references for ss42780946 & ss460400? I've looked in what I thought to be obvious places, but can't find them in dbSNP.

Here is how you find the references:

1. Go to the dbSNP home page, and enter ss42780946 in the text box located in the "Search by ID on all Assemblies" section located just below the announcement section near the top of the page, and select "NCBI Assay ID(ss#)" from the drop down menu to the right of the text box where you entered ss42780946. Click the "search" button.
2. The resulting submitted SNP report will appear. Scroll down this report until you get to the "Individual Genotype Batch" section where you'll see "View citation details". Click on the numeral "1" link just to the right of these words.

3. The resulting [publication detail page](#) does not indicate the actual reference, but if you click on the number that follows the word "PMID" near the top of the page, you will get to the actual publication reference.

I'm sorry the link to the reference is not clear, but I will mention it to the SNP development team, and perhaps they can work out something that will make it easier to see. (8/8/07)

Finding a Gene Name associated with a refSNP (rs) number

How do I retrieve the gene name associated with an refSNP (rs) number I have?

1. Go to the [dbSNP home page](#), and go to the first search section called "Search by IDs on All Assemblies".
2. Type your rs number into the search box (for example, rs11922879).
3. Click on the box to the right of the search box to activate a drop down menu. Select "Reference cluster ID (rs#)".
4. Click on the "search" button.
5. A [refSNP Cluster report](#) for the rs number you entered will appear.
6. Scroll down to the "Geneview" section of the page; it is the third section of the report.
7. If you look directly below the section header, you will find the following statement: "GeneView via analysis of contig annotation: CAV3 caveolin 3." Caveolin 3 (CAV3) is the gene name associated with this refSNP.
8. If you click on "[CAV3](#)" in the sentence mentioned in the previous step, you will see the Entrez Gene short summary of the CAV3 gene.
9. If you want more information than the Entrez Gene summary provides, scroll down a little further in the Geneview section, and you will see the mRNA IDs used in the gene model. Click on one of these, and you will go to a Entrez Nucleotide response. Click on the number representing the mRNA you selected, and you will get another Entrez nucleotide response page. Click on the mRNA ID link presented, and you will see a full report for that mRNA, which include references, authors, and remarks from the mRNA submitter.(4/3/07)

Finding Gene IDs associated with a refSNP (rs) number

Is there an FTP file or a tool (e.g. e-link) that can be used to find the corresponding Gene IDs for a batch of SNPs?

You can get the corresponding gene IDs for refSNP (rs) numbers using [eLink](#), or you can get the information from the SNPContigLocusId table, which can be found in the [organism_data directory](#)(this link takes you to the directory for human) for your organism of interest.

Here is a [description](#) of the columns in the SNPContigLocusId table. Please note that the "locus_id" column of this table stores the NCBI gene ID. (10/10/07)

Finding the Genetic Position of a SNP

How do I get the genetic position of a SNP (i.e. centimorgan [cM]) displayed on top of the SNP's physical position(bp)?

dbSNP contains only the SNP's physical position.(9/18/07)

Finding Strain Information for Submitted SNPs

Where can I obtain available strain information for Rat SNPs in db SNP?

You'll need to look at several data sources to get strain information:

The best data source for strain information is genotype data. As of this date, a small subset of rat SNPs in dbSNP has genotype data. Right now, all of the rat genotype data is being submitted by a group in Netherlands with the handle: [FGG_NIOB](#).

Since most rat SNPs in dbSNP do not have genotype data available, you'll have to check strain information by going to the "method" description of each SNP or to the population information submitted with the SNP (the assay section of the Submitted SNPDetail Report for [ss16346124](#) shows the strain used in the experiment was SHRSP).

1. Another way to find strain information is to view submissions by batch using the following steps:
2. Go to the [dbSNP home page](#)
3. Scroll down and select "New Batches" from the Submission Information Section
4. When you get to the "[View Submissions by Batch](#)" query page, select "rat_10116: Rattus norvegicus" from the first drop-down menu, the variation submission type you want from the second drop-down menu (for the purpose of this example, I used "Variation Submissions SNP"), and click the "search batch" button.
5. The [resulting page](#) will list all the batches of Rat data submitted to dbSNP
6. If you click on the second link (RAT_COMPUTATIONAL_CELERA), you will get the [batch submission page](#) for that batch submission.
7. Click on the words "Rat Strains" in the population section of the page to go to [a list of the strains](#) used in the experiment.

Many rat SNP records do not have any strain information associated with them since the information was never submitted. For these SNPs, you might be able to get the strain information from the original publication associated with the SNP, should such a publication exist. To find out if a publication is linked to a SNP:

1. Go to the [Submitted SNP Detail page](#) for a particular ss number (in this example I'm using [ss52089183](#))
2. If there is a publication cited, it will be in the "submitter" section of the report.

As you can see, obtaining rat strain information is not particularly easy in dbSNP. As this information is of import to the rat research community, we will encourage submitters to submit genotype data and strain information and will try to make this information easier to access on dbSNP. As a start, I have generated a rat strain [report](#) using strain information extracted from genotype data and method data (see above). **(07/01/08)**

Finding CpG Island Data

Where can I get a graphic that contains sequence with CpG islands highlighted or bolded?

You can use Mapviewer to display and view CpG island locations, but at the present time, you will not be able to view CpG islands at the sequence level. To view SNP and CpG island locations:

1. Go to [Mapviewer](#) . Using the drop-down menus at the top, select on organism and gene symbol you wish to use in your search, and click on the "Go" button. For this example, I will use human and BRCA1.

2. Once the search for of your organism and gene symbol is complete, you will be taken to the “[Genome View](#)” page for your search. Scroll down the page until you see a list of “Map Elements” on the right. Select the Map Element of interest to you.
3. You will now be taken to a [MapView](#) page showing the Map Element you selected. Click on the “Maps & Options” button at the upper right-hand side of the page, to go to a window that will allow you to add maps and change display options.
4. To view SNPs and CpG islands, select “CpG Island” from the menu in the left side of the window, and click on “Add”. Then scroll down the list of maps, select “Variation” and click “Add”. At this point, you can add or remove other maps of interest, select which map will be the Master map, or change a number of other map options.

There is [online documentation](#) regarding CpG islands that you can review.

Other online CpG island viewers can be found by searching Google using the terms: "cpg island graphic viewer". (3/22/05)

Finding Flanking Sequence Data

I'm using the Sequenom iPLEX system for genotyping, and need 200bp of 5' and 3' flanking sequence for rs41296860. How do I find this much flanking sequence?

Currently, dbSNP does not have data for long flanks as you request, so you will have to use the following steps to find the

rs41296860 flanking sequence you need:

1. Go to the [Integrated Maps](#) section of the refSNP cluster report for rs41296860.
2. Click on NT_021877.18. This will take you to an [Entrez Nucleotide report](#) showing rs41296860 and its flanking sequence on the genomic contig
3. Since the variation's position spans the contig (NT_021877.18) between 467055 and 467056, all you have to do is subtract or add the number of bases you want from the variation positions, and type the difference/sum in the range: “from” and “to” boxes (e.g. $467055 - 100 = 466955$ and $467056 + 100 = 467156$ for a flank of 100bp on either side of the variation).
4. Click the “Refresh” button to see the fasta sequence in the desired length.

(04/04/08)

I am interested in retrieving flanking sequences in the forward orientation for a list of b126 SNPs. How do I do this?

A refSNP (rs) flanking sequence is simply the flanking sequence of the longest submitted SNP (ss) in the refSNP cluster. The ss with the longest flanking sequence is called the "refSNP exemplar". If a refSNP cluster gets a new ss member added after build 126 and this new ss has flanking sequence that is longer than the flanking sequences of the existing ss in the cluster, then the new ss becomes the refSNP cluster's exemplar, its flanking sequence is adjusted for orientation, and it will be used as the rs cluster's flanking sequence in the next build. Since the new ss will, in most cases, align at the same position as the rs, the flanking sequence difference should be small. I am therefore curious why you would need the rs flanking sequence for build 126. Have you noticed a significant difference (other than length) between rs flanks in different builds?

In general, dbSNP does not keep old build data due to data size issues and the complexity of tracking assembly changes between builds. However, if you have a local copy of dbSNP, you can access the rs flanking sequence for a particular build since dbSNP keeps the flanking sequences of all submitted SNPs. If you do not have a local copy of dbSNP that you can query, give us a list of the rs numbers in question, we can pull the data for you. (11/20/07)

Does dbSNP store flanking sequence for a given refSNP?

We don't store flanking sequence for a given refSNP (rs), but the flanking sequence we use for it is simple to get. We just use the sequence of the member-submitted SNP (ss) that has the longest flank. If the ss with the longest flank is in reverse orientation with the rs, we reverse the ss flank.

Finding SNPs that Lead to Premature Stop Codons

What file do I need to download and parse to identify all SNPs that lead to a premature stop codon?

Use the SNPContigLocusId file. (10/17/07)

Finding Proximal (Neighboring) SNPs

How do I identify proximal SNPs located around a primary SNP?

You can find SNPs that are proximal to a given SNP by using the "Neighbor SNP" link located in the Integrated Maps section of the refSNP Cluster Report. For example, you can find the SNPs proximal to [rs2515644](#) by scrolling down the cluster report until you came to the "Integrated Maps" section, which is located approximately half way down the page. Looking to the right, you'll see the "Neighbor SNP" column. Click on the blue "View" links located in the column to view neighboring SNPs. (10/13/06)

Finding Wildtype Amino Acid

I am working with a local copy of dbSNP and have extracted all the SNPs which affect amino-acid changes. How do I determine which amino acid is wildtype?

There is no designation of a "wild-type" allele since the occurrence of the common allele is population specific. You will have to check the frequency data over each specific population to infer whether the allele is wild type or a mutant in a specific population. Take a look at the description for the [SubPopAllele](#) table. (03/25/08)

Finding Functional (Synonymous/Non-synonymous, etc.) Information for a SNP

Definition of Functional Class

I do not understand the meaning of the "Function" category or any of its components (e.g. synonymous, contig reference, nonsynonymous)?

The NCBI handbook has [documentation](#) that should address your questions. Below is an excerpt from that document that should address your question:

"...[fxn-class] defines variation functional classes. We base class on the relationship between a variation and any local gene features. When a variation is near a transcript or in a transcript interval but not in the coding region, then we define the functional class by the position of the variation relative to the structure of the aligned transcript. In other words, a variation may be near a gene (locus region), in a UTR (mrna-utr), in an intron (intron), or in a splice site (splice site). If the variation is in a coding region, then the functional class of the variation depends on how each allele may affect the translated peptide sequence.

Typically, one allele of a variation will be the same as the contig (contig reference), and the other allele will be either a synonymous change or a nonsynonymous change. In some cases, one allele will be a synonymous change, and the other allele will be a nonsynonymous change. If any allele is a nonsynonymous change, then

the variation is classified as a nonsynonymous variation. Otherwise, the variation is classified as a synonymous variation.

The Four Basic Outcomes When a Variation Is in Coding Sequence

1. The allele is the same as the contig (contig reference) and hence causes no change to the translated sequence.
2. The allele, when substituted for the reference sequence, yields a new codon that encodes the same amino acid. This is termed a synonymous substitution.
3. The allele, when substituted for the reference sequence, yields a new codon that encodes a different amino acid. This is termed a nonsynonymous substitution.
4. A problem with the annotated coding region feature prohibits conceptual translation. In this case, we note the variation class as coding, based solely on position.

Because functional classification is defined by positional and sequence parameters, two facts emerge: (a) if a gene has multiple transcripts because of alternative splicing, then a variation may have several different functional relationships to the gene; and (b) if multiple genes are densely packed in a contig region, then a variation at a single location in the genome may have multiple, potentially different, relationships to its local gene neighbors.” (3/14/05)

I noticed in the documentation where "coding nonsynonymous" is mentioned as a SNP function class, yet, I don't think I've ever seen it used. Instead I see the term "missense".

In ftp reports for older builds, we did use the term “nonsynonymous” to as a function class. Then about a year ago, the functional class code became more sophisticated so that it could describe the type of nonsynonymous change the SNP caused: “missense”, “nonsense” or “frameshift”.

So “missense”, “nonsense”, and “frameshift” are types of nonsynonymous change a variation can cause. (07/15/08)

Finding the Functional Class of a SNP

When I search for rs3093737 using the “Search by IDs on all assemblies” on the dbSNP home page, how do I find the functional class of this SNP (i.e. is it located in an intron)?

1. Scroll down the dbSNP home page until you come to the section marked “Batch”.
2. Look for the sub-section marked “Enter list”, and click on the “Reference SNP ID (rs)” link within this subsection.
3. When the Batch query entry page appears, scroll down until you come to the section marked “Enter RS Numbers” and enter the refSNP number or numbers in the text box for the SNP(s) you are interested in.
4. Scroll down below the “Enter RS Numbers” until you come to the “Select Result Format” page.
5. Click on the text box containing the entry “ASN.1” to activate a drop-down menu. Select “FLATFILE” from the menu and click on the “submit” button.
6. The resulting FLATFILE would normally have the information you are looking for, but as rs3093737 doesn't map to any known gene on the current human genome build (human build 36.2), its location is unknown. (5/1/07)

Finding Non-Synonymous SNPs or information about Non-Synonymous SNPs

Has anyone used dbSNP to determine the median number of nonsynonymous SNPs per gene and the most nonsynonymous SNPs in a gene?

There have been a number of studies that examined nonsynonymous SNPs in the human genome.

Try [searching PubMed](#) using “nonsynonymous”, “SNP” and “genome” or other search terms as key words. (08/28/08)

How can I generate a list of non-synonymous coding polymorphisms with a minor allele frequency of at least 10% for 2000 genes identified either by HUGO gene names or by names like "AA125825"?

You'll have to write a program using [eutils](#) programming utilities:

1. Use eSearch to perform the search and parse out the refSNP number. Look at the following [example](#), which shows a search for SNPs in the human LPL gene (txid9606).
2. Use eFetch to retrieve SNP report with the refSNP number from step 1.
3. You'll have to put your query after the term parameter and remove the LPL gene symbol, since I included it only as an [example](#).
4. To limit your search to non-synonymous SNPs include the filter term "coding nonsynonymous"[Function Class] in the eSearch query. (03/25/08)

What is the best way to extract all refSNPs (regardless of species) whose functional class is "coding nonsynonymous"

You can use eUtils to search Entrez SNP programmatically and retrieve a small subset of SNP records or download the SNPContigLocusID tables from the dbSNP FTP site and search for "coding nonsynonymous" SNP with `fxn_class = 4`.

There is a NCBI [short course](#) in building customized data pipelines using Entrez Programming Utilities (eutils) that you might be interested in, and a [eSearch example](#) is also available.

Here is an [example](#) of the search you would like to try as performed online.

The column description for the SNPContigLocusID table is located in the dbSNP Data Dictionary, and the [human SNPContigLocusID](#) table (to get you started) is located in the [organism_data](#) sub-directory of the human database directory. To retrieve the SNPContigLocusID tables for other organisms, just navigate to the parallel subdirectories for the other organisms. (01/11/08)

How do I search for non-synonymous SNPs in your database?

Conduct your search on Entrez SNP using the term “all[sb]” to retrieve all SNPs, then click on the “Limits” tab near the top of the page, and choose the “coding nonsynonymous” filter, under the “Function class” category. (10/30/08)

Is there any way I can query all of dbSNP to extract all known human SNPs for all C/T SNPs that cause non-synonymous changes in the coding region?

You can only query for all C/T SNPs that cause non-synonymous changes in the coding region by querying Entrez SNP:

1. Start on the [Entrez SNP home page](#)
2. Type "Human" (don't include the quotes) in the text box at the top of the page, and click on the “Go” box.
3. Now, click on the “Limits” tab located just a line below the Search text box to generate a form that contains many limit options.
4. Go to the , the “Functional Class” section, of the limits page and click on the box next to the words “coding nonsynonymous”.

5. Scroll down to the “Observed Alleles” section of the limits page and click the box beside “Y” (to set the limit for C/T). check box, and click on the “Go” box located next to the search text box at the top of the page. The results for this search include 14404 SNPs. (5/3/05)

Finding Genes that Do Not Contain Non-Synonymous SNPs

Is there a way of using dbSNP to determine which genes contain SNPs but do not contain non-synonymous SNPs?

You can search for genes without SNPs in the human genome, but here is no easy search for genes that have SNPs but no non-synonymous SNPs. One option is to download the SnpGeneReport from the dbSNP FTP site ([the human genome_report sub-directory](#)) and filter the data. (12/27/06)

Finding Functional Information for SNPs in Coding Regions

How do I locate synonymous, non-synonymous, and frequency information for SNPs that occur within the coding region of each entry in the RefSeq database?

All SNPs are mapped to a RefSeq sequence. You can search for them using [Entrez SNP](#) and the following query:

1. Enter “all[sb]” into the Search box.
2. Select the following limits: Homo sapiens, coding nonsynon, coding synonymous, and validated by frequency.
3. Click Go.
4. Select the dbSNP Batch Report Display.
5. Select genotypeReport as the result and enter your email address. The result will be sent to you by email.

Finding Functional Data for a Variant Protein Form

How do I determine the enzymatic activity of a catalase SNP in relation to the normal enzyme (i.e. is it 20% of normal, 40% of normal, etc.)?

dbSNP currently does not have functional data for the variant form of the protein. Try searching [PubMed](#) for published information.

This sort of information is difficult to computationally mine from the literature for millions of SNPs. Therefore, dbSNP will be instituting an online user annotation tool for the variation community to contribute functional data and references for SNPs in dbSNP. The annotation tool should be available sometime in the last quarter of 2006. (5/22/06)

Downloading Functional Information for a List of SNPs

How do I download the functional (synonymous/non-synonymous/non-coding) information for a list of multiple RefSNPs?

You can use [Entrez SNP](#) to download the XML, ASN1, or Flat file (FLT) reports, each of which will show SNP functional class using a list of SNP ids (rs#). There are also programming utilities ([eUtils](#)) available to automate this process. The XML, ASN1, or Flat file reports for all SNPs in dbSNP are also available on the dbSNP FTP site. (2/14/05)

Finding Hardy Weinburg Probabilities

Where can I download population diversity data such as Hardy Weinburg Probabilities via FTP for all of the SNPs described in dbSNP?

The information you are looking for is located in the “genotype” file for the particular organism you are interested in. For example, if you are interested in population diversity data for human, go to the dbSNP FTP site, select the [organisms](#) directory, then select the [human](#) directory, and then select the [genotype](#) file. The format of these files is described [online](#). (10/5/06)

Finding Hap-Tagged SNPs

refSNP cluster reports used to have a link to “haplotype tagged SNP” data, but now I can’t find it. Where is it?

We do not currently have “Haplotype tagged SNP” data and therefore have removed the 'Haplotype tagged' search filters from our web pages. We anticipate that we may get “Haplotype tagged SNP” data in the future, and will make them publicly available at that time. (02/27/08)

How do I download a text file or excel file that contains SNPs rs numbers (5kb up and down stream), SNP locations (coding, intron, UTR, etc.), minor allele frequency (for Caucasian) and whether or not the SNP is HapMap tagged for our genes of interest?

Sorry, we don't have a data in the format that you requested. You can get the data as .xml (see instructions below) and parse out the fields of interest.

1. Search Entrez Gene for your [genes of interest](#).
2. To the right of each gene displayed you will see the word “Link”. Click on this word and select “SNP” from the drop-down menu that appears.
3. 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 “XML” as the display type.
4. Click the drop-down menu “Send to” toggle located to the right of the “Display” toggle to see a list of destination options. Select “File” and save the result to your computer.

You can also look at the information using other reports you can select from the display drop-down menu (e.g. FASTA, Flat File, etc.) and aggregate the data from them. (6/21/05)

Finding Specific Heterozygosity Data

How do I get a list of SNPs and their heterozygosity reports in table format for the human SPP2 gene?

1. Search Entrez SNP using the [Gene Name] field.
2. Click on the tab marked “Human”, which is located in the second set of tabs at the top of the page.
3. Just above the “Human” tab you will find a menu of display options. Click on the blue arrow located in the “Display” text box to activate the drop-down menu. Select the “Chromosome Report” option.
4. Once the Chromosome Report option has been selected, you will be taken to the chromosome report format for your search results. If the average heterozygosity is available for your SNPs of interest, they will be displayed under the column entitled “avg het”, which is located toward the right side of the page. (3/24/06)

How do I find validated SNPs located in coding sequence that have heterozygosity over a certain threshold, say 30%?

1. Make a list of gene IDs (the gene ID for the example below is 4023) and look them up on [Entrez Gene](#).
2. Upload the [list of gene IDs](#) on this page. Click retrieve to show the gene results.
3. Select SNP links and click the Display button to see the results.
4. Click on Limits at located at the top of the page, and select your filters (i.e., validated, coding, or heterozygosity).

Finding Orientation Data

I have a huge list of human rs numbers, how do I get a list of their alleles in forward orientation?

This information is available in the [SNPContigLoc](#) and [ContigInfo](#) tables.

SNPContigLoc contains the rs allele on the contig, while ContigInfo contains information about the contig position and orientation on the genome(chromosome), so from these two tables you should be able to derive the genomic allele orientation.

You will find descriptions for the [SNPContigLoc](#) and [ContigInfo](#) tables using the [dbSNP Database Dictionary](#).(11/14/08)

I'm working with a local copy of dbSNP and need to know where I can find information that relates the orientation of a gene with the orientation of a SNP on that gene.

[SNPContigLoc](#) stores rs to Contig orientation and [ContigExon](#) stores mRNA to contig orientation.

Please see the “[Forward vs. Reverse strand Orientation](#)” section in the Mapping Process section of this archive. (05/08/08)

Finding Organism Ploidy

Which dbSNP table/field contains organism ploidy?

dbSNP does not have a table for ploidy. To find the number of unique chromosomes in a particular species, go to the summary page and click on the organism to get to the NCBI taxonomy page. The number of chromosomes for a particular organism is found in the genome information section located in the middle of the page.

Finding Records with OMIM Data/Links

How do I find all the human SNPs that have OMIM links?

There are two ways to access SNPs with Omim links; the easiest way is the following:

1. Go to [Entrez SNP](#), and select the grey “Limits” tab located just below the search text box at the top of the page.
2. Once you are on the Limits page, select “Homo sapiens” from the “Organism” limits box.
3. Scroll down to the “Annotation” limits box and select “OMIM”
4. Scroll back up to the top of the page and click the “Go” button.

If you would like a report that contains a list of all known SNPs for human that have OMIM links, use the [The OmimVarLocusIdSNP](#) table, which is located in your organism’s `organism_data` directory on the dbSNP FTP site. (9/26/07)

How do I get the title of an OMIM variation id?

We don't store OMIM titles in dbSNP. You'll have to look it up in OMIM's [morbidmap table](#). (9/15/05)

When I search for human SNPs and limit the results to chromosome 22, with non-synonymous coding and OMIM links, I got 346 records without OMIM links.

Your original query was (("Homo sapiens"[Organism] AND ("snp omim"[Filter] OR "snp structure"[Filter])) AND "coding nonsynonymous"[Function class]) AND 22[CHR]).

These request limits specify that the results must have OMIM, **or** must have structure; the results therefore, can have either OMIM or have structure — they don't have to have both OMIM and structure.

If you change the query to: (("Homo sapiens"[Organism] AND ("snp omim"[Filter] AND "snp structure"[Filter])) AND "coding nonsynonymous"[Function class]) AND 22[CHR]), your results will include both OMIM and structure. (3/29/06)

What dbSNP report format will provide both a SNP and its specific OMIM ID number?

The easiest way you can get all the SNPs with OMIM links is from Entrez SNP:

1. Go to the [Entrez SNP](#) site.
2. Click on the grey "Limits" tab near the top of the page (just beneath the text search boxes).
3. Select the organism you are interested in from the organism list located at the top left of the page.
4. Scroll down the page almost to the bottom, until you find the list of "Annotation" limits. Select "OMIM"
5. Press the "Go" button located at the top of the page next to the empty text search box, and you will receive a list of your organism's SNPs with OMIM annotation.

You can also get those SNPs with an OMIM ID number by downloading from the dbSNP FTP site: the [OvimVarLocusIdSNP table](#) contains the information you need for your organism of interest (human, in this case). This table is located in your organism's organism_data directory on the dbSNP FTP site.

Column definitions for this table are as follows:

Column Description

- 1 omim_id.
- 2 The locus id the SNP is on
- 3 omim variation id.
- 4 locus symbol
- 5 Amino acid using the contig reference allele.
- 6 Amino acid position in the protein.
- 7 Amino acid of the snp variance.
- 8 var class (used for internal dbSNP processing)
- 9 snp_id (rs#)

Below is an extract from the OvimVarLocusIdSNP table, showing data arranged in columns from left to right in the order mentioned above:

```
100650 217 0001 ALDH2 E 487 K 1 671
102560 71 0003 ACTG1 P 332 A 1 11549200
102574 89 0001 ACTN3 R 577 * 1 1815739
102680 118 0001 ADD1 G 460 W 1 4961
102770 270 0001 AMPD1 Q 12 * 1 17602729
103720 125 0002 ADH2 R 369 C 1 2066702
```

(9/27/07)

Finding Reference Alleles

How do I get the reference allele (ie the allele on the reference contig) for each refSNP in human build 128?

The alleles at the SNP contig position are in table SNPContigLoc.

Specifically for b128, it is in file b128_SNPContigLoc_36_2 which is located in the [organism_data subdirectory](#) for human. The [column descriptions](#) for SNPContigLoc are located in dbSNP's data dictionary. (02/26/08)

Finding Records for a Race or Specific Population

How do I find out what ethnic groups were sampled for human SNPs housed in dbSNP?

Most of the genotype data in dbSNP are from HapMap, and their samples are from four populations. Information about these populations are available at the [International HapMap Project Site](#). (01/03/08)

How do I search for known SNPs in several genes that occur only in the European/Caucasian population?

Although dbSNP does not have a classification for race and ethnic group, you can search on [Entrez SNP](#) for the gene and limit the subset to population class EUROPE. Enter the gene name or term in the search box, click Limits, and check the box for EUROPE under limit by Population Class.

How do I search for SNPs based on a set of criteria that may include race, allele frequency, etc.?

Try your search using [Entrez SNP](#). It has search fields that are available, as well as some examples.

Filter your results using the [limits](#) found in Entrez SNP.

dbSNP is not allowed to classify SNP data based on racial or ethnic information, but you can filter or search SNP data using the field population class, which is based on geographic location.

Finding Sample Size Data

How do I find the number of samples (i.e. people) you've sequenced for a particular polymorphism of human maspin (serpin b5)?

dbSNP does not generate the SNP data. dbSNP is a depository for data submitted from hundreds of research groups. Each SNP may have a different sample size from another SNP. Search for rs2289519 using [Entrez SNP](#). The results page for this search shows a row of colored buttons below rs2289519 that represent links. Click the pink "GeneView" button. This will take you to the "SNP linked to Gene" page, which shows that P176S corresponds to refSNP rs2289519. The [total sample size](#) for rs2289519 is located in the Population Diversity Section of the refSNP report. (6/1/06)

Finding Tab-Delimited Reports

How do I get a tab-delimited report of all mouse SNPs from dbSNP that would show refSNP id and fxn_class?

You can get the refSNP id and fxn_class from the tab-delimited bcp file mmSNPContigLocusId.bcp. mmSNPContigLocusId.bcp can be found in a file called mmNPContigLocusId.bcp.gz, which is located at the SNP FTP site. You can access the SNP FTP site the from dbSNP homepage sidebar. The definition of fxn_class is located in a file called SnpFunctionCode.bcp.gz, and the table columns are defined in dbSNP_main_table.sql.gz, which is located in the [shared Schema](#).

How can I download a tab-delimited file with Population Diversity information for hundreds of SNPs?

There isn't a tab delimited report format containing the data that you requested. You'll have to upload your list of SNPs to the [batch query service](#) to get the XML or ASN report and parse out the data you want.

Finding Validation Data

How do I get a flat file that contains the validation and strain information for all of the mouse strains in dbSNP?

There are two steps:

1. Go to [Entrez SNP](#).
2. Click on Limits and choose *Mus musculus* in the organism section; choose validation status in the Validation section. If you are interested in several specific strains, you could enter the strain name in the search box to narrow the result.

Finding out why a SNP has been Withdrawn

Why has rs11568324 been deleted from dbSNP? We, and other groups have genotyped this SNP, and although the minor allele frequency is very low (0.7%), we consistently found this SNP.

The rs11568324 cluster contained two submitted SNPs (ss#), both of which were from the same submitter. This submitter also submitted a "withdraw" request on September 7, 2006 for the SNPs in question.

The submitter has since re-analyzed their SNP data and has resubmitted their SNPs to us. The submitter indicated that their newly submitted SNPs may include some previously withdrawn SNPs. We are in the process of mapping the new SNPs from this submitter, and if any of these SNPs map to the position that rs11568324 used to map to, they will be assigned to that cluster. (11/21/06)

I searched db SNP several months ago for SNPs in SLC22A2 and its splicing variations, and found non-synonymous SNPs. I searched again today, and found that many of these SNPs have been deleted from dbSNP. Why?

The SNPs mentioned in your question were originally submitted by PHARMGKB (to see the submitter of a ss#, click on the ss#, which is located in the "NCBI Assay ID" column of the "Submitter Records" section of the refSNP report page). PHARMGKB has withdrawn 11740 of their submitted SNPs (ss#) for error corrections and will resubmit the data early next year (2007). PHARMGKB did not indicate the percentage of corrected SNPs that will have flanking sequence changes. (12/19/06)