



## Locating SNPs in a Gene or in Genes

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

### Locating SNPs in a Specific Gene

#### How do I get a table report of all the SNPs associated with MAP3K7?

dbSNP has a report called “GeneView” which will provide a table view of the SNPs associated with a particular gene. I will show you two ways to access the dbSNP GeneView page:

To access the SNP Geneview page from the dbSNP Home page:

1. Go to the [dbSNP Home page](#).
2. Type MAP3K7 into the right-hand text box at the top of the page, and click “GO”.
3. The results page will be [displayed](#). For the purpose of this example, I selected the grey “Human” tab at the top of the results section to display human MAP3K7 SNPs.
4. Select the purple “GeneView” link which is located just below any refSNP entry of interest.
5. This will take you to the [default SNP Geneview page](#), which lists the cSNPs associated with MAP3K7.
6. At the bottom of the “Gene Model” section, you’ll see a blue bar that has a number of display options. Select the button next to “in gene region” and click the “Refresh” button, located immediately to the right.
7. You will now see a [display](#) of all the SNPs associated with MAP3K7.

To access the SNP Geneview page from Entrez Gene:

1. Go to the [Entrez Homepage](#)
2. Select the [Entrez Gene search page](#)
3. Type MAP3K7 into the right-hand text box at the top of the page, and click “GO”.
4. Click on the MAP3K7 link located at the top of the result entry of your choice (I chose the first result, which happened to be Homo sapiens). This will take you to the [Entrez Gene entry for Homo sapiens MAP3K7](#).
5. On the right hand side of the page, there is a “Links” column. Select the SNP: Geneview link.
6. This will take you to the [default SNP Geneview page](#), which lists the cSNPs associated with MAP3K7.
7. At the bottom of the “Gene Model” section, you’ll see a blue bar that has a number of display options. Select the button next to “in gene region” and click the “Refresh” button, located immediately to the right.
8. You will now see a [display](#) of all the SNPs associated with MAP3K7.

(08/11/08)

#### How do I search for polymorphisms in a specific gene?

You could either use Entrez SNP or Entrez Gene as a starting point for your search. To look for SNPs in human epidermal growth factor, for example, using Entrez SNP, do the following:

Go to [Entrez SNP](#)

Type the following search terms in the text box at the top of the page: **[Gene Description] AND [Gene Description] AND [Gene Description] AND [Organism]**. Remember to separate each term of the gene name by the Boolean operator “AND”. For example, if the SNPs you are looking for are located in the human Epidermal Growth Factor Receptor gene, your search in Entrez SNP would look like this: **Epidermal AND Growth AND Factor AND human**

Click on the first refSNP (rs number) in the result list to get the refSNP page for that rs number, which contains more specific information about that specific refSNP.

To look for a SNP in human epidermal growth factor, for example, using Entrez Gene, do the following:

1. Go to [Entrez Gene](#).
2. Enter the search terms in the text box at the top of the page linked by the Boolean operator AND, followed by the filter: “gene snp”[filter] to retrieve genes that contain SNPs (note: be sure to include the quotation marks in the filter text). Using the human epidermal growth factor example, the search would look like this: epidermal AND growth AND factor AND “homo sapiens” AND “gene SNP” [filter].
3. Click the “Go” button to the right of the text box.
4. On the results page, to the far right of each record is the word “Links” in blue text — click on it to activate a drop-down menu of NCBI resources that are linked to this record.
5. Select “GeneView in dbSNP”, which is located toward the bottom of the menu.
6. You will get the dbSNP report for that locus record.(04/05/06)

### **How do I obtain a list of all the known variants (SNPs, indels, etc.) within a gene, and within the 10 kb immediately 5’ and immediately 3’ of the same gene?**

The instructions below will show you how to query the gene database to get the 5’ and 3’ chromosome positions for the gene and then query dbSNP for 10 kb on either side of those positions:

1. Get the position from the graphic display on [Entrez Gene](#).
2. Query using the gene symbol/name or gene/locus ID.
3. Display the result as graphics.
4. Record the chromosome number, the 5’ gene position, and the 3’ gene position.
5. Subtract 10 K from the 5’ gene position and add 10 K to the 3’ gene position.
6. Query Entrez SNP using chromosome number and the newly calculated gene positions.
7. Example: 8[CHR] AND 19807051:19835042[CHRPOS]

### **How do I search for all the SNPs located in a particular gene?**

You can search using the gene symbol on Entrez SNP and by clicking the graphic L link for a record to get an overview all SNPs in the same gene.

[Here's an example URL](#).

### **How do I find out if there are microsatellite variations in the VDR gene?**

1. Go to [Entrez SNP](#), and place the text “VDR” in the search box near the top of the page.
2. Click on the “limits” tab, which is located directly below the text search box near the top of the page.
3. Scroll down to the “organisms” limits section and choose "homo sapiens"
4. Scroll down further to “the “SNP class” limits section and choose "microsat"
5. Scroll up to the top of the page and push the “Go” button. (9/6/06)

## Locating SNPs in a Gene in a Graphical Format

**Where is the resource that allows you to actually see in a graphic how many SNPs are known to lie within a particular gene, and their physical coordinates?**

The resource you refer to is the sequence viewer graphical display of a gene region on the human genome. We had to discontinue this resource because the NCBI sequence viewer itself fails when the sequence under consideration is very large (contig sized).

You can still get a sequence view of the mRNA record using the seqview button on the Entrez SNP display, but this won't show intronic SNPs. There may be other inconsistencies even among coding SNPs since the position of SNPs on mRNAs is taken from a different mapping pipeline than the one used to compute genomic locations.

For now, you can search Entrez SNP using the gene symbol or gene id and click the 'GeneView' icon on any of the displayed rs records:

1. Go to Entrez SNP and enter the search terms "LPL AND Human" (without the quotation marks) into the text box at the top of the page, and click on The "Go" button.
2. You will get a page back that contains 158 results to your query.

You can also use Entrez SNP to get a rough text representation of a SNP position in a gene by entering the gene symbol in the search box at the top of the page. When you get the result page, select "chromosome report" from the "display" drop-down menu and "Chromosome Base Position" from the "sort by" drop-down menu. (6/3/05)

## Locating a SNP in a Gene Using a Gene Name

**Where can I search SNP using gene names? Can I perform these searches in batch form by submitting a text list?**

Entrez SNP will allow you to search dbSNP using gene names. EntrezSNP will not, however, allow you to conduct a batch search using gene names. To search using multiple gene names, you can use the operator OR in your query: for example, "LPL OR BRCA1 OR EF2".

**How do I search for polymorphisms in a specific gene?**

You could either use Entrez SNP or Entrez Gene as a starting point for your search. To look for SNPs in human epidermal growth factor, for example, using Entrez SNP, do the following:

Go to [Entrez SNP](#)

Type the following search terms in the text box at the top of the page: **[Gene Description] AND [Gene Description] AND [Gene Description] AND [Organism]**. Remember to separate each term of the gene name by the Boolean operator "AND". For example, if the SNPs you are looking for are located in the human Epidermal Growth Factor Receptor gene, your search in Entrez SNP would look like this: **Epidermal AND Growth AND Factor AND human**

Click on the first refSNP (rs number) in the result list to get the refSNP page for that rs number, which contains more specific information about that specific refSNP.

To look for a SNP in human epidermal growth factor, for example, using Entrez Gene, do the following:

1. Go to [Entrez Gene](#).
2. Enter the search terms in the text box at the top of the page linked by the Boolean operator AND, followed by the filter: "gene snp"[filter] to retrieve genes that contain SNPs (note: be sure to include

the quotation marks in the filter text). Using the human epidermal growth factor example, the search would look like this: epidermal AND growth AND factor AND "homo sapiens" AND "gene SNP" [filter].

3. Click the "Go" button to the right of the text box.
4. On the results page, to the far right of each record is the word "Links" in blue text — click on it to activate a drop-down menu of NCBI resources that are linked to this record.
5. Select "GeneView in dbSNP", which is located toward the bottom of the menu.
6. You will get the dbSNP report for that locus record.(04/05/06)

### **How do I find the genomic coordinates of a gene if I know the gene name, for example, scnn1a?**

Search for the gene name on [Entrez Gene](#) .

Click on the blue text that says "Links" located on the far to the right of the scnn1a result to activate a drop down menu, and select "GeneView in dbSNP".(6/7/06)

### **Is there a bad link to the SNP records for NRP2 (human neuropilin 2)? When I key RP2 into the Entrez SNP search box, it takes me to NELL2, a completely different gene.**

By keying only "RP2" into the search box, you are telling Entrez to search not just the gene name field but all fields that contain "RP2"; so you end up with 1465 records, most of which do not pertain to your gene of interest. If you key in "RP2[Gene Name]" into the Entrez SNP search box, then Entrez SNP will retrieve only those SNP records that contain RP2 in the "Gene Name" field ( 352 records).

### **Why is it that when I searched dbSNP for "vimentin" I got no hits, but if I used "vim" instead of "vimentin", I got 112 hits?**

dbSNP doesn't index the full gene name. You can search dbSNP using the gene symbol or gene ID, or you can search [Entrez Gene](#) using the gene name and then click the word "Links" to the left of the entry of interest to see a dropdown menu of NCBI resources with data related to the entry. Select "SNP" to see SNPs related to the entry you selected. (2/9/07)

## **Locating SNPs in a Named Gene**

### **Why is there no SNP data for the factor VIII gene on the X chromosome? There are known polymorphisms in the gene.**

Since dbSNP is a catalog of variations submitted mainly by our users, if a known polymorphism is not in dbSNP, it just means that Factor VIII has yet to be submitted. dbSNP adds new data daily, so perhaps SNPs for Factor VIII will be submitted in the near future. Anyone can submit to dbSNP using our [online instructions](#). (9/25/07)

### **I accidentally generated a list of SNPs for CFTR. How so I do this for other genes?**

1. Go to [Entrez Gene](#).
2. Type "ABO[gene] AND Human[organism]" (with no quotes) in the text box located at the top of the page and press the "Go" button.
3. Look at the far right hand side of the page. You will see an alphabetical list of links. Scroll down until you see the entry "SNP: Geneview". Click on this entry.

**Please Note:** The default format for GeneView in dbSNP shows SNPs located in coding regions only. To see all of the SNPs in the gene, go to the grey bar just above the second section on the page [marked "gene model (contig mRNA transcript)"]. Select the radio button in the grey bar marked "in gene region" and then click the "refresh" button located at the right end of the grey bar to see all the SNPs in the gene. (3/22/07)

**I looked up the enzyme "catalase" on Entrez Gene and found the associated SNP page. It states that there are 240 entries for Homo Sapiens. Are there 240 SNP variations that have currently been identified?**

There are 240 unique reference SNPs in dbSNP for the human CAT gene. There may be more identified in the literature that have yet to be submitted to dbSNP. (5/22/06)

**How do I search for SNPs located in the various cytochrome P450 enzymes?**

1. There are many ways to use the NCBI web pages to view SNPs located in the cytochrome P450 enzymes. Below are the steps for one way to do this search for a specific gene — CYP27B1 — in the cytochrome P450 enzyme family.
2. Since you have a list of gene products (the enzymes), and you wish to find SNPs associated with those products, I would suggest starting from [Entrez gene](#).
3. Type "CYP27B1" (without quotes) in the text box at the top of the Entrez Gene page and click the "Go" button. You will get the Entrez Gene entry for the CYP27B1 gene.
4. Located on the far right side of the web display there is text that says "links". Click on the word "links" to get a dropdown menu of information options. Click on the "GeneView in dbSNP" option. This will take you to a page listing the SNPs in and around the CYP27B1 gene.
5. Repeat steps 2 and 3 above for each gene. (2/8/06)

**How do I determine the number of SNPs located within the human tubulin2 gene (TUBB2)? The accession number is NP\_006079.**

If you want to locate all the SNPs in and around the TUBB2 gene, you can do so by using one of the following two methods:

Method 1:

Go to [Single Nucleotide Polymorphism](#), enter "TUBB2" into the search bar at the top of the page, and dbSNP will retrieve eight SNPs.

Method 2:

Go to the [refSNP by Locus](#) page. Here, you can see that five SNPs are in the gene region of TUBB2, one SNP is in the genomic region of TUBB2, and GenBank annotation records show two additional SNPs. To have dbSNP show all eight SNPs, click the Send button. This action will send the list to batch query, which will locate and list all eight rs numbers.

**How do I find the nucleotide position information for the Leu432Val SNP in CYP1B1? I found its rs number (rs1056836) in Entrez Gene.**

1. Go to the [SNP homepage](#).
2. Locate the SEARCH by IDs section of the page, type rs1056836 in the text search box, and click on the Search button.
3. You'll be taken to the rs report page for rs1056836. Go to the [GeneView section](#).
4. Click on the U03688 link. This will take you to the Entrez nucleotide [report](#) for this sequence:.
5. Select GenBank from the drop-down list of display options located at the top of the page, and you'll be given the GenBank record for this sequence. Scroll a third of the way down the page to locate the CDS (coding sequence) annotation. Make a note of the start codon for the sequence. Repeat steps 4 and 5 for NM\_000104.

You'll find that the start codon for the CYP1B1 gene is located at position 347 for U03688 and at position 373 for NM\_000104: CDS 347..1978 for U03688, CDS 373..2004 for NM\_000104.

Translated into English, the above records indicate that the RefSeq (U03688) starts 26 bp upstream (the 5' UTR) of the GenBank mRNA (NM\_000104).

You can use the RefSeq record above to determine the position of SNP rs1056836 relative to the start codon by doing the following: Subtract 1 from the start (codon) position,  $347 - 1 = 346$ . Now, subtract 346 from the position of the SNP,  $1640 - 346 = 1294$  exons.

If you use the GenBank mRNA record to determine the position of SNP rs1056836 relative to the start codon,  $373 - 1 = 372$ , then  $1666 - 372 = 1294$  exons. Divide the number of exons by 3 to determine the codon position of the SNP:  $1294/3 = 431.3$ . It looks like the SNP is in the first position of the 432 codon.

Of course, if any of these sequences were to contain an insertion or deletion with respect to one another, the final positions would be different. It is therefore best to reference a sequence and position rather than just a position.

Nucleotide positions are enumerated starting at 1 for each record shown on the dbSNP page; mRNA and RefSeq positions skip introns.

### 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)

## Locating a Specific Variation Type in a Named Gene

### How do I find out if there are microsatellite variations in the VDR gene?

1. Go to [Entrez SNP](#), and place the text "VDR" in the search box near the top of the page.
2. Click on the "limits" tab, which is located directly below the text search box near the top of the page.
3. Scroll down to the "organisms" limits section and choose "homo sapiens"
4. Scroll down further to "the "SNP class" limits section and choose "microsat"
5. Scroll up to the top of the page and push the "Go" button. (9/6/06)

## Locating a SNP in a Gene using an Amino Acid Variation

### I know that there is a Glu3405-->Gln change in a gene. How do I find the nucleotide variation (G>C) responsible for it in dbSNP?

You can search for amino acid variations using HGVS nomenclature and NCBI protein accessions. Please see the protein examples on the "[Human Variation: Search, Annotate, Submit](#)" page (scroll down to the bottom). (09/02/08)

## Locating SNPs in Specific Regions within a Gene

### How do I find the sequence position of a SNP on a gene?

As far as I know NCBI doesn't have a canonical set of "gene records" that would provide a coordinate system for SNP alignment to genes. NG\_ records are curated gene regions, but they do not exist for all genes annotated on the human genome. In any event, we make no systematic effort to map our SNPs to this set apart from NCBI genome alignment.



dbSNP maintains a table called ContigExon which provides the start and stop positions of each exon in contig coordinates for each mRNA refseq defined in a particular NCBI genome build. We use these coordinates to create the exon placement graphic on our [Geneview display](#).

You can obtain the same canonical list of exon boundary information in chromosome coordinates from an NCBI genomes ascii file. (4/20/05)

## Locating SNPs in the UTR Region of a Gene

**I would like to find human SNPs located in the 3' utr region of the AGTR1 gene, but the standard dbSNP searches do not show SNPs in just the 3' UTR.**

Right now, the easiest way to find the known SNPs in just the 3'UTR region of the AGTR1 gene is to do the following:

1. Go to [Entrez Gene](#) and key in "AGTR1" in the search box at the top of the page, and click "Go"
2. On the resulting page, click on the AGTR1 symbol above the record you are interested in.
3. On the resulting record, find the "Links" section located on the right-hand side of the page, and click on the words "SNP: Geneview".
4. On the resulting record, scroll down to the bottom of the "Gene Model" section, where you will find a series of radio buttons in a grey bar. Click the button marked "in gene region" and click the "refresh" button.

You can also try searching the SNPContigLocusId table. For human, you would use b128\_SNPContigLocusId\_36\_2, which is located in the [human subdirectory](#) of the organism\_data directory on the dbSNP FTP site. The table contains gene symbols, so you'll be able to distinguish SNPs associated with certain genes. Look for SNPs that have Fxn\_class = 53, which means that the SNP is in 3' UTR. You can find a more detailed description about the table in the dbSNP [data dictionary](#).

We are currently working on an annotation process for SNPs located in genes with more detailed functional class information. Among other things, this process will note whether a particular SNP is located in the 3'UTR or the 5'UTR. Once this annotation process is implemented, a user will be able to query for SNPs located in the 3'UTR region of a given gene. We are close to finishing this; it may be in place in a month or two.

Another tool currently under development that may be of use to you in the future is a SNP genome workbench plug in which will make it easy for users to query by gene and gene functional details. (11/02/07)

## Locating SNPs in Multiple Copy Genes

**What part of the SNP database contains the known SNPs for the human 28S rRNA nuclear-encoded gene? I'm having problems orienting myself with multiple copy genes.**

1. Go to the [Entrez Gene](#) site.
2. Copy and paste the following query into the textbox at the top of the Entrez Gene page: "genotype rrna"[Properties] AND Human[orgn]

This query returns 10 records, 2 of which are Mitochondrial. Record LOC100008589 seems to be the 28S, however, it is a provisional record.

Your query has prompted us to evaluate how we can better display variations in non-protein encoding genes in dbSNP. The NCBI Gene group is looking into the hueristics of annotating these to the genome. (3/23/07)

## Locating SNPs in Genes from a Specific Population

**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.

## Locating Genes that Contain a Large Number of SNPs

**Is there any way of identifying genes that contain a large number of SNPs or retrieving SNP data by using the number of SNPs/gene as a query term?**

dbSNP doesn't have a web service for this kind of search. You will have to download the [SNPContigLocusId table](#), located in the organism\_data directory for your organism (human in this case), and query it using SQL. (4/20/05)

## Locating SNP Information for a Large Number of Genes

**How do I get SNP information (refSNP number, contexts, SNP type, etc.) for 1000 genes?**

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](#). (2005)

## Locating SNPs in Promoters

**How do I locate SNPs in promoter regions?**

SNPs are divided into the functional classes, shown below. This list does not include promoter regions. We'll look into adding promoter regions in the future.

SNP functional classes:

- locus region
- intron
- coding nonsynon
- mrna utr
- coding synon
- reference
- exception
- splice site

**How do I find the SNP location for the G/C polymorphism in the IL6 promoter located at base 174 and the G/C polymorphism found in the following sequence: ttgtgtcttgc(G/C)atgctaaa ggacgtcaca ttgcacaatc ttaataaggt ttccaatcag?**

SNP locations are stored as chromosome positions in dbSNP. Use the sequences flanking the variation of interest to MegaBlast against the chromosome database; the results will include the chromosome number and



the position where the SNP of interest is located. Once you have the chromosome number and the position, you can use them to query Entrez SNP. Please see the example below:

Example:

1. Use [Megablast](#) to get the chromosome position(s).
2. Use the chromosome number and position(s) to query Entrez SNP. [This link](#) shows “no items found”. This means that the SNP has not yet been submitted to dbSNP.
3. If the variation had been submitted to dbSNP, you would have found a “submitted SNP” accession number (ss#), or “reference SNP” number (rs#) cited in the literature. [Here is a link](#) to other SNPs located in IL6 that you might be interested in.

### How do I find the starting and ending positions of a gene?

To obtain gene starting and ending positions on a contig, follow the instructions below once you get your SNP search results:

1. Select Display, then Gene Links on the Entrez SNP page (toward the top).
2. Select the gene of interest on the Entrez Gene.
3. Select Display, then Gene Table on the Entrez Gene page (toward the top).
4. Make a note of the starting and ending positions on the graphic and use them to search on Entrez SNP using CTPOS.

If you know the position of your regulatory region, then search with the contig position [CTPOS] using the upstream offset from the starting position of your gene. You can also use the chromosome position [CHRPOS] in conjunction with the chromosome number field [CHR].

## Locating SNPs in Exons/Introns

### How do I get the position of a SNP when it is located in an intron?

You can see SNPs located in intronic regions in the "GeneView" page by clicking on the "in gene region" radio button and then click the “refresh” button, both of which are located in the blue bar just below the mRNA alignment section near the top of the page.

For example, if you look at the Geneview page for [LPL](#), you’ll see 15cSNPs in the LPL gene, but if you click on the on the "in gene region" radio button, and then the “refresh” button, you’ll see [all the SNPs](#) (234 SNPs — including many intronic SNPs) in dbSNP that are associated with this gene.

(11/20/08)

### Which SNP report format will tell me if a SNP falls in an exon or an intron?

Use the flatfile format.(05/27/08)

### After locating SNPs within a gene, is there a way to scan for SNPs in exons vs. introns, and missense vs. silent polymorphisms, without examining each SNP record?

Yes, use Entrez SNP. Below is an example of this quick scanning technique using the CLOCK gene in humans:

1. Go to the [Entrez SNP](#) search page or the [dbSNP homepage](#).
2. Enter CLOCK[GENE] AND HUMAN[ORGN] in the Search box and click Go.
3. 99 SNPs will be returned in the search result. There is a graphic bar under each snp\_id in the search result. Within that bar, there will always be an L, which stands for “locus”. If the L is blue, then the SNP is in the locus (our search only returned SNPs at a gene locus).

4. Just to the right of the L is a T, which stands for “transcript”. If the T is blue, then the SNP is also in the transcript. To the right of the T is a C, which stands for “coding region”. If the C is any color other than wash-out white, then the SNP is also in a coding region. If the C is green, then the SNP will produce a non-synonymous change in the protein. If the C is red, then it is a synonymous change.
5. You can restrict a query to a particular type by checking the appropriate boxes on the Limits form found in the function class section of the page. Access the Limits form from the Limits link, directly under the search form.
6. You can ask for the union of those SNPs that either code synonymous or are located in introns to be returned by checking the corresponding boxes. There are 42 such SNPs for the CLOCK gene. Or, you might only be interested in non-synonymous changes for the CLOCK gene; in this case, set the coding non-synonymous box on the Limits form, and then you’ll find that only two SNPs are returned, rs3762836 and rs1056478.

### **How do I search for verified SNPs that are contained on expressed material (5’ UTR, 3’ UTR, or exons) and are polymorphic between B6 and BALB/c PLUS mice?**

Go to the Search Mouse SNP between strains tool and [limit the results](#) on Entrez using the function class and genotype filter.

### **How do I determine if there are any SNPs in the first exon/intron in the factor ix human gene?**

Please follow the search instructions below.

1. Search Entrez SNP [using the terms "factor ix AND human"](#).
2. Click on the L graphic link for the first SNP on the page; this will take you to the [summary page](#) for SNP → LocusLink.

### **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).

### **How do I search for information regarding the number of SNPs found in coding regions within a gene using the gi, NM, or XM ID numbers?**

Go to the [organism\\_data](#) directory located in your organism’s database and click on SNPContigLocusId.bcp.gz. Below is an ASCII image of the database table containing all of our snp-to-gene information as gathered from the NCBI assembly processing. See below for column headings.

You will probably only want to keep rows where fxn\_class is 3, 4, or 8.

Here are the columns. The example is the first NON\_SYNON encountered from the top:

Columns	Example	COMMENT
snp_id	47	snp_id
contig_acc	NT_007819	contig accession
contig_ver	14	contig version
asn_from	10876794	contig position
asn_to	10876794	contig position
locus_id	23249	locus link identifier
locus_symbol	KIAA0960	gene name

mrna_acc	XM_371877	mrna accession
mrna_ver	2	mrna version
protein_acc	XP_371877	protein accession
protein_ver	2	protein version
fxn_class	4	2,3,4,8 are coding, 6=intron, 5=utr
reading_frame	1	
allele	G	allele as found on mrna_acc
residue	D	residue as found on protein_acc
aa_position	514	position of residue I protein_acc
build_id	34_3	Genome build; assembly context is a property of contig

## Locating Exon Start/End positions

### How do I find exon starting and ending positions within a particular gene?

We have a file with the exon starting and ending positions within a contig in table ContigExon. You can get the file ContigExon.bcp.gz in the [organism\\_data](#) (link goes to human\_9606 database as example) directory found within your organism's database.

### How do I find the starting and ending positions of a gene?

To obtain gene starting and ending positions on a contig, follow the instructions below once you get your SNP search results:

1. Select Display, then Gene Links on the Entrez SNP page (toward the top).
2. Select the gene of interest on the Entrez Gene.
3. Select Display, then Gene Table on the Entrez Gene page (toward the top).
4. Make a note of the starting and ending positions on the graphic and use them to search on Entrez SNP using CTPOS.

If you know the position of your regulatory region, then search with the contig position [CTPOS] using the upstream offset from the starting position of your gene. You can also use the chromosome position [CHRPOS] in conjunction with the chromosome number field [CHR].

## Locating SNPs in Stop Codons

### How do I format a search for all the stop codons in dbSNP?

I do not know of a way to query either the dbSNP homepage or Entrez SNP to get that information. If you have a local copy of the dbSNP database, you can do the following query:

Select "snp\_id,locus\_symbol, protein\_acc FROM SNPContigLocusId, WHERE residue = '\*' to get all the SNP- protein relationships (that we can identify) for organism=human, where the SNP is in the stop codon. If you don't have a local copy of the dbSNP database, you can parse the same information from the SNPContigLocusId table located in your organism's [organism\\_data](#) (this link is to the human\_9606 directory as an example). (see columns 1,7,10).

## Locating SNPs in Intergenic Regions

### I'd like to know on what basis intergenic SNPs are assigned to genes in dbSNP. Is it based on a certain distance that the SNP is from the gene?

Intergenic SNPs are assigned to genes in dbSNP based on distance.

They are assigned to a gene if they are within 0.5kb 3' to a gene, or within 2kb 5' to a gene.

A good way to see if a particular SNP is associated with a gene is to look at the SnpFunctionCode for the SNP of interest. To do this, first look in the [shared\\_data](#) directory for your organism on the dbSNP [ftp site](#) and download SnpFunctionCode.bcp.gz, which defines the codes used for function class. Then go to the [organism\\_data directory](#)(this link takes you to the directory for human) for your organism of interest, and download SNPContigLocusId.bcp.gz, which contains the snp\_id and the functional class code for each SNP. You can find a [description](#) of the columns in the SNPContigLocusId table online.

Looking at the codes defined in SnpFunctionCode.bcp.gz, you'll see that if the function code is 13 for the SNP, then the functional class is "NearGene-3", which means that the SNP is within 0.5kb 3' to a gene. If the function code for the SNP is 15, then the functional class is "NearGene-5", which means that the SNP is within 2kb 5' to a gene. (10/3/07)

### **How do I obtain SNPs located in human intergenic regions?**

Search using the limit strategy below:

1. Enter "all[sb]" to select all SNPs.
2. Click on Limits, then check and enter the following limits: CBID Range from 1 to 119, chromosome 1, Homo sapiens, snp[SnpClass], Weight 1. Result for query #1 = 393170.
3. Click on Limits again, then check and enter the following limits: CBID Range from 1 to 119, chromosome 1, coding nonsynon, reference, exception, intron, coding synonymous, locus region, mrna utr, splice site, Homo sapiens, snp[SnpClass], Weight 1. Result for query #2 = 180392.
4. Click on History, which shows you the queries and results.
5. Uncheck the Limits box.
6. Use the search numbers located next to each query to do a Boolean query. For this example, assume your queries above have been assigned the search numbers "#1" and "#2". Enter in the search box "#1 NOT #2" without the quotes. Result for query #3 = 212778.