Locating and Analyzing SNPs Associated with Disease Causing Mutations

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

dbSNP Locus Specific Database (LSDB)/Clinical Resources

What Locus Specific Database (LSDB)/Clinical SNP submission resources are available on dbSNP?

dbSNP created two web forms dbSNP to facilitate submission of LSDB/Clinical SNPs, both of these forms require the user to register themselves as a data submitter (submitter HANDLE) so that they can be credited with their contributions, and to describe the submitted variant according to the HGVS standard (i.e. relative to a reference sequence and version.

The HumanVariation: Search, Annotate, Submit site allows users to:

- Enter the HGVS description of a single variant and ascertain if the variant is already in the database by determining if the entry matches an existing rs number
- Contribute novel information, MIM numbers, PubMed UIDs, and free text to new and existing variants
- Review their submission before making it available to users

The HumanVariation: Annotation and Submit Batch Data site allows the user to provide all the information that they can in the single submission form (see above), except that they can do so for a large number (Batch) of records. In addition, this resource allows the user to

- Provide identifiers for each variant
- Provide URLs back to the originating LSDB database, both globally and for each variant ID
- Enter the number of independent observations of each variant.

(**06/17/09**)

What search terms do I use in Entrez SNP to find clinically associated SNPs that have links with a Locus Specific Database (LSDB)?

Use the search term "snp_gene_clin"[filter] in Entrez SNP; this filter is also included in as an annotation limit on the Entrez SNP "limits" page. Here is an example of the results page for such a search. For access to the LSDB links, click the VarView button for a specific result to go to the VarView page, which will have links to LSDBs and other resources located in the upper right of the page.(**06/16/09**)

Locating SNPs Associated with Disease Causing Mutations

Is it possible to narrow the search on dbSNP to mutations that have been published in the literature as disease-causing mutations?

Yes, you can do so by following the directions below. **Please Note**: the use of the word "mutation" is being phased out in dbSNP, and will be replaced by the term "Clinical/LSDB variation":

1. Go to the Entrez SNP home page.
2. Place your search term(s) in the search box at the top of the page.
3. Select the Entrez Limits tab.
4. When the "limits" page appears, scroll down to the "Annotation" box and Select "Clinical/LSDB variations" and/or "omim"
5. Click the "Go" button at the top of the page.

Originally, the great majority of data in dbSNP was collected and defined as variations simply using sets of co-aligned genomic or DNA sequences. Because this process typically had little to no focus on disease condition, only about 250 records in dbSNP were successfully associated with phenotype-causing variations or a clinical outcome in OMIM.

Starting in the Spring of 2008, however, dbSNP began accepting submissions of Clinical/LSDB variations as well as annotations to existing variations (including phenotype) on the Human Variation: Search, Annotate, Submit site (for single submissions) as well as on the Human Variation: Annotate and Submit Batch Data site (for multiple submissions). Currently there are a total of 1266 records in dbSNP that were submitted as Clinical/LSDB variations (select "Clinical/LSDB variation" in the Entrez SNP limits page and click "GO" with out entering a search term in the Entrez search box), and 1134 records submitted as Clinical/LSDB variations that also have OMIM links (select "Clinical/LSDB variation" and "OMIM" in the Entrez SNP limits page and click "GO" with out entering a search term in the Entrez search box).

Those SNPs with clinical association(s) will have a red "VarView" (Variation Viewer) link in the "allele" section at the upper right of the refSNP cluster report. Clicking the link will take you to the Variation Viewer Report for the gene in which the SNP is found.

We expect that the number of Clinical/LSDB variation records in dbSNP will grow rapidly as more users discover dbSNP's resources for submitting them. (**07/09/08**)

We are trying to find data on drug resistance-causing mutations. Are you aware of a database that lists submissions by protein, organism, point mutation, and drug?

We are not aware of a database that lists specifically the data you requested, but you could try searching OMIM for mutations affecting drug resistance.

Although the data you require are not presented strictly as a list, you may have to read the original publication listed on OMIM to get the details. You could also try searching Google for "drug resistance mutation database". Google search

What is the difference between rs10735810 and rs2228570, and which should I use for breast cancer research?

According to dbSNP, rs10735810 is a SNP in the exon of the VDR gene that causes a non-synonomous change in its amino acid and has been genotyped by many submitters, while rs2228570 has no unique mapping.

Based on the above information, I hope you can make you own choice as to which SNP to study. (**9/6/06**)

Analysis of SNP Data Associated with a Specific Disease

Could you suggest a statistical program to analyze data from a SNP that may be associated with a certain kind of heart failure?

There are many different statistical packages that could perform the type of analysis that you need to do. A comprehensive list of all types of genetic software is available.

Rather than suggesting a particular statistical package for the evaluation of allelic association with a complex phenotype such as heart disease, I would suggest that you involve a statistician for the actual analysis because there are pitfalls to conducting such research without proper statistical knowledge. Please take a look at the following review: *The Lancet: Problems of reporting genetic associations with complex outcomes* [PubMed].

I searched dbSNP for "diabetes mellitus" and got 2585 SNPs. How do I determine the number of genes on which these SNPs are located?

Go to the Locus Information Query page. Choose human from the organism list and type in "diabetes mellitus" in the query box. The search will take you to LocusLink, which will show a count of 66 loci; this is the number of genes relating to "diabetes mellitus" that have SNPs.

The 2585 SNPs in your search result are the number of hits to the LocusLink description "diabetes mellitus". Very few dbSNP records have phenotype and disease descriptions. That is why you don't see "diabetes mellitus" mentioned in dbSNP records of these SNPs.