



Locating ENCODE (Encyclopedia Of DNA Elements) Project SNPs

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

How do I query dbSNP to get the ENCODE SNPs?

dbSNP received ENCODE region SNPs in 2004 from a submitter with the handle "BROAD".

Currently, there are two queries you can use to get ENCODE SNPs:

Query Option A — conduct a "Method" search:

1. Go to the [dbSNP Home page](#)
2. Click on "Search" in the left blue side bar. This will release a set of options. Select "Method", which will take you to the "Search/View Method Detail" page.
3. At the top of the page you will see a number of "Search by" options. Select "Submitter method id" and then select "contains" from the list just above the text box.
4. Type "ENCODE" (without the quotes) into the text box, and click the Search button.
5. You will find two "BROAD" method IDs. Click on either of the two Method IDs to see method descriptions and a list of submission batches that used the particular method you selected.
6. Click on any of the submitter Batch IDs listed to see a list of submitted SNPs numbers submitted in that batch for the ENCODE project.

Query Option B — Retrieve the ENCODE FTP files:

1. Go to the dbSNP FTP site, and find the `/organisms/human_9606/database/organism_data/` directory
2. The following two files should provide the information you are looking for:
[Encode_NT2NC.bcp.gz](#)
[b125_SNPencode_35_1.bcp.gz](#)
3. The schema file for the ENCODE table ([ENCODE_table_DDL.txt](#)) is located at the dbSNP FTP site in the `/organisms/human_9606/misc/` directory.

Please note that the ENCODE region data is on NCBI build 35. As of this date, there is no update on the current build(36). (10/14/08)

Where can I find SNPs from the ENCODE pilot re-sequence project?

All of the ENCODE re-sequenced SNPs are located in the `/ENCODE_resequence_submitted_snp.txt.gz` file in the `/organisms/human_9606/misc/` directory of the dbSNP FTP site. At the top of the file you will find a description of the data and column definitions. The re-sequenced SNPs were submitted by BROAD and BCM (Baylor):

The 13545 ENCODE re-sequenced SNPs submitted by BROAD all share Submitted Method ID that has the word "ENCODE" in it, so you can search for it using the "Method" search described in another FAQ* in this section.

The 16734 ENCODE re-sequenced SNPs from BCM (Baylor) all share the same publication that has "ENCODE" in the title, so you can find these SNPs by doing the following:

1. Go to the [dbSNP Home page](#)
2. Select "Search" from the left blue side bar.
3. Select "publication" from the resulting list of choices.
4. At the top of the resulting ([Search/View Publication Detail](#)) page you will see a number of "Search by" options. Select "Publication Title" and then select "contains" from the list just above the text box.
5. Type "ENCODE" (without the quotes) into the text box, and click the Search button.
6. The [results](#) show "ENCODE Re-sequencing" as a title. If you click on the "ENCODE Re-sequencing" title on this page, the [resulting page](#) shows all the batches submitted to dbSNP that cited the "ENCODE Re-sequencing" paper.

(10/14/08)