



dbSNP Build Updates and Schema Changes

Created: August 8, 2005; Updated: June 15, 2010.

I noticed that b127 SNPs are no longer associated with function class codes 1,2,5, and 7. Does dbSNP no longer use these function class codes?

As of build 127, function codes 1, 5 and 7 have been modified into two digit codes that will more precisely indicate the location of a SNP. The two digit codes have function codes 1, 5 or 7 as the first digit, each of these numbers keeping its original meaning, and 3 or 5 as the second digit, indicating whether the SNP is 3' or 5' to the region of interest. So the new function codes are as follows:

Function code 13: "nearGene-3"

Where:

1=locus region

3= SNP is 3' to and 0.5kb away from gene

Function code 15: "nearGene-5"

Where:

1=locus region

5= SNP is 5' to and 2kb away from gene

Function code 53: "UTR-3"

Where:

5= UTR (untranslated region)

3= SNP located in the 3' untranslated region

Function code 55: "UTR-5"

Where:

5= UTR (untranslated region)

5(as the second digit)= SNP located in the 5' untranslated region

Function code 73: "splice-3"

Where:

7=splice site

3=3' acceptor dinucleotide

Function code 75: "splice-5"

Where:

7=splice site

5=5' donor dinucleotide

Function code 2, however, was retired permanently as of b127, as it identified a SNP as being in the coding region of a gene, but that other details about its location were unknown. Since mapping and annotation have improved dramatically since function code 2 was defined, it is no longer used. (4/30/07)

I noticed that the coding non-synonymous function code has been subdivided. Can you list and define the subdivisions?

Function Code 41: “Nonsense” (coding nonsynonymous)

changes to the Stop codon

Function Code 42: “Missense” (coding nonsynonymous)

alters codon to make an altered amino acid in protein product

Function Code 44: “Frameshift” (coding nonsynonymous)

indel SNP causing frameshift

You can find up-to-date function codes and their definitions in the [SnpFunctionCode.bcp.gz](#) table located in the /shared_data directory of the dbSNP FTP site. (10/27/08)

Does each build represent schema changes or data updates?

Each build represents both data updates and schema changes.

When can we expect schema documentation for build 125 — it is very confusing trying to understand all the schema changes without any documentation.

Try looking at the documentation in the [Database Dictionary](#). The introduction to the dictionary states:

“Table/column descriptions should be updated whenever there are database changes. This task takes time since dbSNP has well over a hundred tables. Instead of waiting for all table/column descriptions to be completed before releasing the data dictionary, we are letting a table/column description go public as soon as it is finished to allow faster user access. Please click on “Contact Us” on the sidebar if the table of your interest still lacks the updated description.”

I regularly check the content of each table, and update descriptions when needed. This is an ongoing task, so thank you for your patience. (10/25/05)

Your web site directed me to the SNP.bcp and SnpValidationCode.bcp tables to find the validation status for each SNP, but the data dictionary and the database schema don’t match up with the fields in SNP.bcp.

The [updated SNP table description](#) can be found online, and the introduction to our [Schema Dictionary](#) states:

“Table/column descriptions should be updated whenever there are database changes. This task takes time since dbSNP has well over a hundred tables. Instead of waiting for all table/column descriptions to be completed before releasing the data dictionary, we are letting a table/column description go public as soon as it is finished to allow faster user access. Please click on “Contact Us” on the sidebar if the table of your interest still lacks the updated description.” (10/25/05)

Would it be possible for you to send me a preview of upcoming changes to the schema for build 125?

The major change in build125 will be for dbSNP to be split into organism-specific databases so as to increase dbSNP performance and to make dbSNP easier to maintain. Splitting dbSNP into organism-specific databases should not affect most users if they are interested in only one organism.

Mapping tables (SNPMapInfo, SNPContigLoc and ContigInfo) will have some changes, but the submission related tables as well as the frequency and genotype related tables have not changed. (6/15/05)