



## The Difference Between the SNP Build and the Genome Build

Created: July 7, 2005; Updated: February 18, 2014.

### What is the difference between the NCBI dbSNP build and the genome build?

Genome builds are a release of assemblies made from contiguous sequences (contigs) assembled in what is thought to be chromosomal order. This allows researchers to have a common scaffold on which they can annotate genomic features in chromosomal coordinates.

The dbSNP database is a collection of user submitted variations (flanking sequence and variable alleles), allele frequencies, and genotypes for these variations. dbSNP "maps" the variations to the genome assemblies, numbers them with submitted SNP or ss numbers, clusters the ss numbers into refSNP or rs numbers, and then performs summary statistics as well as many other tasks.

The dbSNP build is a release of the dbSNP database and is usually synchronized with the subsequent genome builds. There may be more than one dbSNP build for each genome build, however. Each genome build (i.e., human genome build 36, dbSNP build 126) consists of one or more assemblies. For example, human genome build 36 has reference and Celera assemblies for all chromosomes. Alternate assemblies may also occur in certain regions, such as across the HLA DRB genes on chromosome 6p. (2/2/06)

### Is there a relationship between the SNP build number and the human genome build number?

Take a look at the [build history](#) of dbSNP; this will show you the dbSNP build numbers and human genome build numbers. (10/18/05)

### I noticed in the build history of dbSNP that build 119 is mapped to human build 34, version 2. What does "version 2" mean?

The NCBI assembly group has recently extended the build ID from a single number (e.g., 34) to a version number (e.g., 34.1). Build 34.2 was the first build to adopt the new numbering system, whereas build 34 was aliased to build 34.1 at that time.

The new build ID convention is that the build number will change when there is a change to the assembly sequence itself, so the build will be given a new version number when gene, SNP, STS, and other feature annotations are updated on unchanged sequence.

Now that the sequencing of the human genome is nearly completed, this sensible revision of the original build schedule allows new information about features to be added asynchronous of the assembly process itself.