



dbSNP and HapMap Data

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

What version of HapMap data was included in dbSNP build 128?

dbSNP build 128 was released on October 24, 2007, and contains HapMap release 21a data. **(08/18/08)**

How do I get the age of the 270 HapMap samples?

dbSNP does not have age information for the HapMap samples. What we do have is the pedigree information for those individuals. Since all 270 samples for the HapMap project have Coriell sample IDs, you could also view information for these samples on the Coriell site. Here is the Coriell Data summary for Coriell ID#:

[HAPMAPPT07](#)

(10/23/06)

How do I determine the hapmap block for rs893584?

We do not have haplotype block information as calculated by the HapMap project, but we do have estimated haplotypes submitted as a joint collaboration with Eleazar Eskins' group at UCSD using all of the genotype data submitted to the dbSNP as of build 125 (including HapMap Phase I genotypes). The XML files containing these estimated haplotypes are located in the dbSNP FTP site, within the [haplotypes folder](#) of the human sub-directory. **(4/20/06)**

Is phasing data for Caucasian Parent-Child trios found in HapMap included in the dbSNP tables?

Phased genotypes for HapMap and all dbSNP submissions are available at this time via the FTP site only. The phasing was performed by Eleazer Eskin's group at UCSD. Please see the new data format located in the [human haplotype directory](#). I would advise you to read the [readme](#) file in this directory for a description of this directory's contents. **(1/13/06)**

Is dbSNP data is regularly updated with the new data from the HapMap project? Does the NCBI SNP data set have any advantage over the HapMap data set?

The SNPs chosen by HapMap originated from dbSNP entries, therefore, HapMap is a SUBSET of dbSNP's refSNP (rs) numbers.. All Phase I genotype data from HapMap has been submitted back to the dbSNP. The Phase II genotype data will be released with dbSNP b126.

The technology employed by the HapMap limits the variations used to true SNPs (i.e. two alleles and one bp) existing in dbSNP at the time submitters designed their assays. Subsequent submissions to dbSNP will therefore not be in HAPMAP. There are also many cases of SNPs typed by HapMap that have been genotyped by other projects which provide additional genotype and allele frequency information.

dbSNP also contains other variations (e.g. SNPs with more than two alleles, indels, MNP's, and microsatellite markers), many of which have genotype and/or allele frequency information. **(1/10/06)**

How do I name population samples?

Look at the [HapMap discussion](#) on naming population samples. (5/17/05)

The HapMap project lists dbSNP as a source for its SNPs, but if this is true, then the dbSNP data in HapMap would not confirm or validate the SNP. It seems like circular logic to me.

The HapMap project does get its SNP sequences from dbSNP, but these SNPs are then sequenced over a set of individual samples by the HapMap project. This is why we specifically set the "HapMap" validation status. Please see more details at the [HapMap site](#).

HapMap Genotype Discrepancies

Why is the genotype data in dbSNP for rs2235961 (Perlegen) contradictory to data derived from the same samples (HapMap-CEU) also housed in dbSNP?

dbSNP does not generate genotype data. All genotype data reported in dbSNP are submitted to us from various laboratories. The [Population Diversity](#) section of the refSNP report for this SNP indicates that there are conflicting genotype calls for this SNP from different submitters. These conflicting calls could be due different methods and differing sample quality. Please contact the individual submitters for further information. (06/09/09)

I've noticed that in a number of SNPs (e.g. rs34950166 and rs35040247) all individuals examined in the 4 HapMap populations are always heterozygous. This seems unlikely.

The refSNPs you mentioned were genotyped in a QA project by PERLEGEN using HapMap samples. For example, if you go to the genotype section of the refSNP cluster report for rs34950166, and click on ss68759579, you will see in the page below that the [genotype data](#) was submitted whose submitter ID (handle) is "GAIN-PERLEGEN-QC".

We have noticed the genotype data quality issues for a number of batches from "GAIN-PERLEGEN-QC" and "GAIN-BROAD-QC", and one of our staff members has been working with these submitters to fix the genotype data. We hope the data will be corrected in time to be released with the next build (B130). (07/17/08)

The genotype results for rs34958084 shows are strange: in the ss48428804 assay, 100% of CEU, HCB and JPT individuals are homozygous for T; but, in the assay ss66405533 assay, 100% of CEU, HCB and JPT individuals are homozygous for C.

You will see as you examine the reports for rs34958084, that the submitter of the genotype data was "GAIN-BROAD-QC".

We have noticed the genotype data quality issues for a number of batches from "GAIN-PERLEGEN-QC" and "GAIN-BROAD-QC", and one of our staff members has been working with these submitters to fix the genotype data. We hope the data will be corrected in time to be released with the next build (B130). (07/17/08)