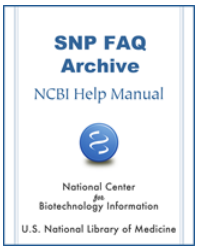
U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: SNP FAQ Archive [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2005-. Discrepancies between Cluster Report data and Other Genome Resources. 2005 Jul 7 [Updated 2014 Feb 18].
Bookshelf URL: https://www.ncbi.nlm.nih.gov/books/



Discrepancies between Cluster Report data and Other Genome Resources

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

Ensembl shows rs3210531 on chromosome 11 strand +1, however, NCBI shows it on chromosome 6 strand -1. Why?

We conduct two different mapping procedures for each SNP. First one places the SNP on contigs, and according to our database, rs3210531 hits chromosome 11 across all assemblies:

snp_id	pos	chr	assembly
3210531	46406925	11	reference
3210531	46598418	11	Celera
3210531	46157042	11	HuRef

An additional hit to chr6 was rejected due to our "absolute hit" strategy. This strategy states that if the mapping process finds a 100% identity hit, then the process will reject hits with lower scores as irrelevant.

Mapping to underlying sequences is a separate process and it produced the following hits for rs3210531:

snp_id	offset	accession	accession_ver	aln_quality
3210531	366	U14972	1	1
3210531	423	BC001955	1	1
3210531	423	BC005012	1	1
3210531	369	BC001032	2	1
3210531	367	BC070235	1	1
3210531	379	BC071946	1	1
3210531	456	BC073799	1	1
3210531	418	NM_001014	3	1

We are working now on a synchronization mechanism for two these processes. **(09/29/08)**

The data for rs13306584 is not the same in the UCSC Genome Browser, Ensembl, or dbSNP. Why is this?

The UCSC Genome Browser and Ensembl haven't yet updated to build 128, which is the current dbSNP build. USCS is updated only to build 126, while Ensembl is updated to build 127. Since the data for rs13306584 has changed between these three builds, the data for the three databases will not be the same until the UCSC Genome Browser and Ensembl update to dbSNP build 128. (**11/23/07**)