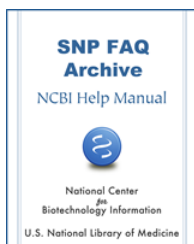




U.S. National Library of Medicine
National Center for Biotechnology Information

NLM Citation: SNP FAQ Archive [Internet]. Bethesda (MD): National Center for Biotechnology Information (US); 2005-. Discrepancies between dbSNP Summary Data and FTP Data. 2005 Jul 7 [Updated 2014 Feb 18].

Bookshelf URL: <https://www.ncbi.nlm.nih.gov/books/>



Discrepancies between dbSNP Summary Data and FTP Data

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

The summary page of a particular build implies that there are 7,231,721 rs numbers. I interpret your FTP files to have 7,238,948, a difference of 7227. Can you resolve this difference?

There are about 7,200 redundant rs numbers in the FTP file, because SNPs in the pseudo-autosomal region (PAR) are reproduced in both the X and Y chromosome files. The introduction of PAR into the NCBI assembly is fairly recent. The rs set in our FTP files used to be unique—not so any more.