

## How to Find the SNP in Real Data

The same SNP will always has the same "rs" name under different dataset (build). But the same SNP may have different position tag under different build. So if you want to extract the SNP by "rs" name, it always OK. But if you want to extract the SNP data by the position, make sure the build and position information is correct.

How to get the SNP position tag (or the "rs" name) for a particular gene?

We may need to use 2 websites: 1. UCSC Genome Bioinformatics: [http:// genome.uscs.edu](http://genome.uscs.edu) 2. NCBI: <http://www.ncbi.nlm.nih.gov>

Suppose we want to search the information for Gene RBJ, and finally extract the SNP data from the Gain data, which is Rel #3, May 10, on NCBI B36 assembly.

1. On the homepage of UCSC Genome Bioinformatics, click the top left "Genome Brower"
2. Choose the correct assembly, here we need to choose "Mar, 2006 (NCBI36/hg18)", which is identical to the information to both the gain data set and HAPMAP
3. In the "gene" box, input RBJ, but there is no match result. The reason is that RBJ may have another name in UCSC database.

To find the correct name for Gene RBJ, we use the second website NCBI. In the homepage of NCBI, choose to search "Gene", then input "RBJ", click "search" and we choose the name from "[homo sapiens]" which means human. For RBJ, it is "DNAJC27", then use this new name to search the information in the UCSU database.

4. Then we jump to a figure, which shows all SNPs that UCSU can give us. However, not all the SNP are useful, to filter out some garbage SNP, we may use the option "SNP 130". An explanation for SNP 130 can be found on the website of UCSC. Then "hide all", click "full" in "Ref Seq Genes" in "Gene and Gene Prediction Tracks" and click "full in SNP(130)" in "variation and Repeats".
5. To get the text format of the SNPs, click "Tables" on the top of the search result, make sure the assembly is correct, then click the "get the output". You will finally get the result