

# XtraSNPlocs.Hsapiens.dbSNP144.GRCh37

April 10, 2016

---

XtraSNPlocs.Hsapiens.dbSNP144.GRCh37

*The XtraSNPlocs.Hsapiens.dbSNP144.GRCh37 package*

---

## Description

Extra SNP locations and alleles for Homo sapiens extracted from NCBI dbSNP Build 144. The source data files used for this package were created by NCBI on May 29-30, 2015, and contain SNPs mapped to reference genome GRCh37.p13.

While the **SNPlocs.Hsapiens.dbSNP144.GRCh37** package contains only molecular variations of class *snp*, this package contains molecular variations of other classes (*in-del*, *heterozygous*, *microsatellite*, *named-locus*, *no-variation*, *mixed*, and *multinucleotide-polymorphism*).

## Details

SNPs from dbSNP were filtered to keep only those satisfying the 3 following criteria:

- The SNP is NOT a single-base substitution (i.e. its class is NOT *snp*) but is a molecular variation that belongs to any other class supported by dbSNP: *in-del*, *heterozygous*, *microsatellite*, *named-locus*, *no-variation*, *mixed*, or *multinucleotide-polymorphism*.
- The SNP is marked as notwithdrawn.
- A *single* location on the reference genome (GRCh37.p13) is reported for the SNP, and this location is on chromosomes 1-22, X, Y, or MT.

## Note

The source data files used for this package are the same as those used for the **SNPlocs.Hsapiens.dbSNP144.GRCh37** package and were created by the dbSNP Development Team at NCBI on May 29-30, 2015.

The SNPs in this package are mapped to reference genome GRCh37.p13. Note that GRCh37.p13 is a patched version of GRCh37 however the patch doesn't alter chromosomes 1-22, X, Y, MT. GRCh37 itself is the same as the hg19 genome from UCSC *except* for the mitochondrion chromosome.

**Author(s)**

H. Pages

**References**SNP Home at NCBI: <http://www.ncbi.nlm.nih.gov/snp>dbSNP Human BUILD 144 announcement: <http://www.ncbi.nlm.nih.gov/mailman/pipermail/dbsnp-announce/2015q2/000163.html>GRCh37.p13 assembly: [http://www.ncbi.nlm.nih.gov/assembly/GCF\\_000001405.25/](http://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.25/)hg19 genome at UCSC: <http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg19>

Note that chromosomes 1-22, X, and Y in hg19 and GRCh37.p13 are the same except that they are named differently (no chr prefix in GRCh37.p13).

**See Also**

- The **SNPlocs.Hsapiens.dbSNP144.GRCh37** package for SNPs of class *snp*.
- **XtraSNPlocs** objects in the **BSgenome** software package for how to access the data stored in this package.
- The **GRanges** class in the **GenomicRanges** package.
- The **VARIANTAnnotation** software package to annotate variants with respect to location and amino acid coding.

**Examples**

```
## -----
## A. BASIC USAGE
## -----
snps <- XtraSNPlocs.Hsapiens.dbSNP144.GRCh37
snpcount(snps)

## Get the location, RefSNP id, and alleles for all "extra SNPs" on
## chromosome 22 and Y:
my_snps1 <- snpsBySeqname(snps, c("ch22", "chY"), c("RefSNP_id", "alleles"))
my_snps1

## Get the location and alleles for some RefSNP ids:
my_rsids <- c("rs367617508", "rs398104919", "rs3831697", "rs372470289",
             "rs141568169", "rs34628976", "rs67551854")
my_snps2 <- snpsById(snps, my_rsids, c("RefSNP_id", "alleles"))
my_snps2

## -----
## B. COMPUTE AND ADD REFERENCE ALLELE AS AN ADDITIONAL METADATA COLUMN
## -----
library(BSgenome.Hsapiens.UCSC.hg19)
genome <- BSgenome.Hsapiens.UCSC.hg19

## Before we can call getSeq(genome, my_snps1), we need to harmonize the
```

```
## seqinfo components of 'genome' and 'my_snps1':
seqlevelsStyle(my_snps1) # dbSNP
seqlevelsStyle(genome) # UCSC
seqlevelsStyle(my_snps1) <- seqlevelsStyle(genome)
genome(my_snps1) <- "hg19"

## Also hg19 and GRCh37.p13 have incompatible chromosome MT so we must
## drop this seqlevel:
seqlevels(my_snps1) <- seqlevelsInUse(my_snps1)

ref_allele1 <- getSeq(genome, my_snps1)
ref_allele1[ref_allele1 == ""] <- "-"
mcols(my_snps1)$ref_allele <- ref_allele1
my_snps1

## -----
## C. COMPARE ALLELES REPORTED BY dbSNP WITH REFERENCE ALLELE
## -----
alleles1 <- mcols(my_snps1)$alleles
alleles1 <- CharacterList(strsplit(alleles1, "/", fixed=TRUE))
disagrees_idx <- which(all(as.character(ref_allele1) != alleles1))
my_snps1[disagrees_idx]
length(disagrees_idx) / length(my_snps1) # 0.003105851
## Conclusion: 0.31% of the "extra SNPs" in dbSNP have reported alleles
## that disagree with the computed reference allele :-/
```

# Index

\*Topic **package**

XtraSNPlocs.Hsapiens.dbSNP144.GRCh37,

[1](#)

GRanges, [2](#)

XtraSNPlocs, [2](#)

XtraSNPlocs.Hsapiens.dbSNP144.GRCh37,

[1](#)

XtraSNPlocs.Hsapiens.dbSNP144.GRCh37-package

(XtraSNPlocs.Hsapiens.dbSNP144.GRCh37),

[1](#)