XtraSNPlocs.Hsapiens.dbSNP141.GRCh38

April 11, 2018

`XtraSNPlocs.Hsapiens.dbSNP141.GRCh38`
The XtraSNPlocs.Hsapiens.dbSNP141.GRCh38 package

Description

Extra SNP locations and alleles for Homo sapiens extracted from NCBI dbSNP Build 141. The source data files used for this package were created by NCBI on May 1st, 2014, and contain SNPs mapped to reference genome GRCh38.

While the **SNPlocs.Hsapiens.dbSNP141.GRCh38** 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 (GRCh38) 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.dbSNP141.GRCh38** package and were created by the dbSNP Development Team at NCBI on May 1st, 2014.

Author(s)

H. Pages

References

SNP Home at NCBI: http://www.ncbi.nlm.nih.gov/snp

dbSNP Human BUILD 141 announcement: http://www.ncbi.nlm.nih.gov/mailman/pipermail/dbsnp-announce/2014q2/000139.html

GRCh38 assembly: http://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.26/

See Also

- XtraSNPlocs in the **BSgenome** software package for how to access the data stored in this package.
- The **VariantAnnotation** software package to annotate variants with respect to location and amino acid coding.

Examples

```
snps <- XtraSNPlocs.Hsapiens.dbSNP141.GRCh38
snpcount(snps)

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

## Compute and add reference allele as an extra metadata column:
library(BSgenome.Hsapiens.NCBI.GRCh38)
genome <- BSgenome.Hsapiens.NCBI.GRCh38
seqlevelsStyle(my_snps1)  # dbSNP
seqlevelsStyle(genome)  # NCBI
seqlevelsStyle(my_snps1) <- seqlevelsStyle(genome)
ref_allele1 <- getSeq(genome, my_snps1)
ref_allele1[ref_allele1==""] <- "-"
mcols(my_snps1)$ref_allele <- ref_allele1
my_snps1

## 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)  # 0.01097903
## Conclusion: more than 1% of the "extra SNPs" in dbSNP have reported
## alleles that disagree with the reference allele :-/
```

Index