SNPlocs.Hsapiens.dbSNP155.GRCh38

August 7, 2024

SNPlocs.Hsapiens.dbSNP155.GRCh38

The SNPlocs.Hsapiens.dbSNP155.GRCh38 package

Description

Human SNP locations and alleles extracted from dbSNP Build 155 and placed on the GRCh38/hg38 assembly

Details

The 949,021,448 SNPs in this package were extracted from the RefSNP JSON files for chromosomes 1-22, X, Y, and MT, located at https://ftp.ncbi.nih.gov/snp/archive/b155/JSON/ (these files were created by NCBI in May 2021). These SNPs are compatible with packages BSgenome.Hsapiens.NCBI.GRCh38 and BSgenome.Hsapiens.UCSC.hg38, that is, they can be “injected” in the BSgenome objects contained in these packages. SNP positions and alleles are reported with respect to the plus strand.

Only SNPs of type snv (single-nucleotide variant a.k.a. single-base substitution) were kept. Other variant types supported by dbSNP are: delins (indel), ins (insertion), del (deletion), and mnv (multiple nucleotide variation). These other variants are NOT included in SNPlocs.Hsapiens.dbSNP155.GRCh38 but are available in the XtraSNPlocs.Hsapiens.dbSNP155.GRCh38 package.

Note

The SNPs in this package can be "injected" in BSgenome.Hsapiens.NCBI.GRCh38 or BSgenome.Hsapiens.UCSC.hg38, and will land at the correct positions.

See ?injectSNPs in the BSgenome software package for more information about the SNP injection mechanism.

Author(s)

H. Pagès
References

The hg38 genome at UCSC (based on GRCh38.p13, as of April 2022, but the UCSC folks could change this in the future and base hg38 on a more recent patch release of GRCh38): http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg38

See Also

- The XtraSNPlocs.Hsapiens.dbSNP155.GRCh38 package for SNPs of type other than snv.
- snpcount in the BSgenome software package for how to access the data stored in this package.
- IUPAC_CODE_MAP in the Biostrings package.
- The GPos class in the GenomicRanges package.
- injectSNPs in the BSgenome software package for SNP injection.
- The VariantAnnotation software package to annotate variants with respect to location and amino acid coding.

Examples

### A. BASIC USAGE

```r
snps <- SNPlocs.Hsapiens.dbSNP155.GRCh38
snpcount(snps)
seqinfo(snps)
```

```r
## Get the positions and alleles of all SNPs on chromosomes 22 and MT:
snpsBySeqname(snps, seqnames=c("22", "MT"))
```

```r
## Get the positions and alleles of all SNPs within some regions:
snpsByOverlaps(snps, GRanges(c("Y:230001-232000", "19:88501-89000")))
```

### B. EXTRACT SNP INFORMATION FOR A SET OF RS IDS

```r
my_rsids <- c("rs2639606", "rs75264089", "rs73396229", "rs55871206",
              "rs10932221", "rs56219727", "rs73709730", "rs55838886",
              "rs3734153", "rs79381275", "rs1516535", "rs74342513")
```

```r
## Note that the first call to snpsById() takes a long time but
## subsequent calls are faster.
my_snps <- snpsById(snps, my_rsids)
my_snps
```
## Translate the IUPAC ambiguity codes used to represent the alleles into nucleotides:
IUPAC_CODE_MAP[mcols(my_snps)$alleles_as_ambig]

## C. INJECTION IN THE REFERENCE GENOME

library(BSgenome.Hsapiens.UCSC.hg38)
ref_genome <- BSgenome.Hsapiens.UCSC.hg38
ref_genome

alt_genome <- injectSNPs(ref_genome, "SNPlocs.Hsapiens.dbSNP155.GRCh38")
alt_genome # note the additional line "with SNPs injected from..."

alphabetFrequency(ref_genome$chr22)
alphabetFrequency(alt_genome$chr22)

## Get the number of nucleotides that were modified by this injection:
neditAt(ref_genome$chr22, alt_genome$chr22) # 12798921
Index

* package
  SNPlocs.Hsapiens.dbSNP155.GRCh38,
  .loadAlleles
    (SNPlocs.Hsapiens.dbSNP155.GRCh38),
  .loadLoc
    (SNPlocs.Hsapiens.dbSNP155.GRCh38),
  BSgenome, 
COMPATIBLE_BSGENOMES
  (SNPlocs.Hsapiens.dbSNP155.GRCh38),
  GPos, 
injectSNPs, 
IUPAC_CODE_MAP, 
snpcount, 
SNPlocs.Hsapiens.dbSNP155.GRCh38, 
SNPlocs.Hsapiens.dbSNP155.GRCh38-package
  (SNPlocs.Hsapiens.dbSNP155.GRCh38),