SNPlocs.Hsapiens.dbSNP155.GRCh37

March 6, 2024

SNPlocs.Hsapiens.dbSNP155.GRCh37

The SNPlocs.Hsapiens.dbSNP155.GRCh37 package

Description

Human SNP locations and alleles extracted from dbSNP Build 155 and placed on the GRCh37/hg19 assembly

Details

The 929,496,192 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 package BSgenome.Hsapiens.UCSC.hg19, that is, they can be "injected" in the BSgenome object contained in this package.

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.GRCh37 but are available in the XtraSNPlocs.Hsapiens.dbSNP155.GRCh37 package.

Note

The SNPs in this package can be "injected" in BSgenome.Hsapiens.UCSC.hg19 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 hg19 genome at UCSC (based on GRCh37.p13): http://genome.ucsc.edu/cgi-bin/hgGateway?db=hg19

See Also

• The XtraSNPlocs.Hsapiens.dbSNP155.GRCh37 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

```r
## A. BASIC USAGE
snps <- SNPlocs.Hsapiens.dbSNP155.GRCh37
snpcount(snps)
seqinfo(snps)

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

## 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
my_rsids <- c("rs2639606", "rs75264089", "rs73396229", "rs55871206",
               "rs10932221", "rs56219727", "rs73709738", "rs55838886",
               "rs3734153", "rs79381275", "rs1516535", "rs74342513")

## 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.hg19)
ref_genome <- BSgenome.Hsapiens.UCSC.hg19
ref_genome

alt_genome <- injectSNPs(ref_genome, "SNPlocs.Hsapiens.dbSNP155.GRCh37")
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) # 11798497
Index

* package
  SNPlocs.Hsapiens.dbSNP155.GRCh37,
    .loadAlleles
      (SNPlocs.Hsapiens.dbSNP155.GRCh37),
    .loadLoc
      (SNPlocs.Hsapiens.dbSNP155.GRCh37),
  BSgenome, 1
COMPARABLE_BSGENOMES
  (SNPlocs.Hsapiens.dbSNP155.GRCh37),
GPos, 2
injectionSNPs, 1, 2
IUPAC_CODE_MAP, 2

snpcount, 2
SNPlocs.Hsapiens.dbSNP155.GRCh37, 1
SNPlocs.Hsapiens.dbSNP155.GRCh37-package
  (SNPlocs.Hsapiens.dbSNP155.GRCh37), 1