PolyPhen.Hsapiens.dbSNP131

November 11, 2026

`PolyPhen.Hsapiens.dbSNP131`
PolyPhen predictions for Homo sapiens dbSNP build 131

Description

Database of PolyPhen predictions for Homo sapiens dbSNP build 131

Details

- Methods : See ?'PolyPhenDb-class' for methods.
- Creation of Database Tables : This package includes PolyPhen-2 predictions for dbSNP build 131 human coding non-synonymous SNPs. The primary table, "ppdata" is composed of pph2_snp131_missense_HumDiv-full.txt and pph2_snp131_missense_HumVar-short.txt (see below for details on file content). The second table, "duplicates" contains the data from snp131_duplicate_rsids.txt formated in a two-column dataframe with all rsids in the first column and a duplicate group number in the second. The duplicate group was created to identify the groupings; it has no other significance.

  Original PolyPhen files were cleaned as follows :
  - Column names of B-fact and H-bonds were renamed as B_fact and H_bonds.
  - Question marks '?' were replaced with NA
  - The 'rsid' column in "ppdata" was created from the 'snp_id' column.
- Source Files :
  - Source : UCSC Genome Browser GRCh37/hg19 assembly annotation, snp131 track
  - Software : PolyPhen-2 v2.0.22r308
  - Databases : UniProtKB/UniRef100 release 2010_07, Jun 15, 2010_07 UCSC PDB snapshot 2010-06-10 DSSP snapshot 2010-06-10 Pfam 24.0 (October 2009)
  - Source Files : pph2_snp131_missense_HumDiv-full.txt PolyPhen-2 summary output including full set of features; HumDiv classifier model pph2_snp131_missense_HumVar-short.txt PolyPhen-2 short summary output including prediction outcome and scores; HumVar classifier model snp131_duplicate_rsids.txt List of duplicate dbSNP rsIDs

- Description : This package contains PolyPhen-2 annotations for 110,940 human missense SNPs; 5,517 of them do not include mutation effect predictions (as indicated by the keyword "unknown" in "prediction" column). Lack of predictions is explained by either insufficient number of sequence homologs found (indicated by 'NA' in the "Nobs" column) or by the variation site falling withing a gapped region of multiple sequence alignment (indicated by '0' in the "Nobs" column).
  The "Comments" column of the summary files contains original hg19 chromosome coordinates and alleles of each missense SNP extracted from the UCSC snp131 track. All alleles listed are on the plus strand of the reference assembly.
  Approximately 13,000 dbSNP reference SNP IDs annotated missense SNPs with identical chromosome position / alleles, thus translating into a same amino acid residue substitution. Only the first one of each set of such duplicate rsIDs is listed in "snp_id" column of the summary files. To aid in mapping other duplicate rsIDs, snp131_duplicate_rsids.txt file is provided which contains one set of duplicate rsIDs per line, with first rsID corresponding to the one listed in the "snp_id" column of summary files.
  There were 3,137 SNPs for which none of the alleles listed matched reference nucleotide at the chromosome position. Such SNPs were considered dubious and excluded from the analysis.

Author(s)

Valerie Obenchain <vobencha@fhcrc.org>

References

PolyPhen Home: http://genetics.bwh.harvard.edu/pph2/dokuwiki/

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. Nat Methods 7(4):248-249 (2010).

Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. Nucleic Acids Res 30(17):3894-3900 (2002).

Sunyaev SR, Eisenhaber F, Rodchenkov IV, Eisenhaber B, Tumanyan VG, Kuznetsov EN. PSIC: profile extraction from sequence alignments with position-specific counts of independent observations. Protein Eng 12(5):387-394 (1999).

See Also

PolyPhenDb-class

see ? `PolyPhenDbColumns` for column descriptions

Examples

```
library(PolyPhen.Hsapiens.dbSNP131)

## metadata
metadata(PolyPhen.Hsapiens.dbSNP131)

## column descriptions found at ?PolyPhenDbColumns
head(keys(PolyPhen.Hsapiens.dbSNP131))
cols(PolyPhen.Hsapiens.dbSNP131)
```

```
## subset on keys and cols
subst <- c("AA1", "AA2", "PREDICTION")
rsids <- c("rs2142947", "rs3026284")
select(PolyPhen.Hsapiens.dbSNP131, keys=rsids, cols=subst)

## retrieve substitution scores
subst <- c("IDPMAX", "IDPSNP", "IDQMIN")
select(PolyPhen.Hsapiens.dbSNP131, keys=rsids, cols=subst)

## retrieve the PolyPhen-2 classifiers
subst <- c("PPH2CLASS", "PPH2PROB", "PPH2FPR", "PPH2TPR", "PPH2FDR")
select(PolyPhen.Hsapiens.dbSNP131, keys=rsids, cols=subst)

## snps that have been reported under multiple rsids
duplicateRSID(PolyPhen.Hsapiens.dbSNP131, c("rs71225486", "rs1063796"))
```

Index

∗ **data**
PolyPhen.Hsapiens.dbSNP131, 1
∗ **package**
PolyPhen.Hsapiens.dbSNP131, 1

PolyPhen.Hsapiens.dbSNP131, 1
PolyPhen.Hsapiens.dbSNP131-package
*(*PolyPhen.Hsapiens.dbSNP131*)*, 1
PolyPhenDb-class, *2*