

SIFT.Hsapiens.dbSNP137

January 19, 2022

SIFT.Hsapiens.dbSNP137

PROVEAN/SIFT predictions for Homo sapiens dbSNP build 137

Description

Database of PROVEAN/SIFT predictions for Homo sapiens dbSNP build 137

Details

The SIFT tool is no longer actively maintained. A few of the original authors have started the PROVEAN (Protein Variation Effect Analyzer) project. PROVEAN is a software tool which predicts whether an amino acid substitution or indel has an impact on the biological function of a protein. PROVEAN is useful for filtering sequence variants to identify nonsynonymous or indel variants that are predicted to be functionally important.

See the web pages for a complete description of the methods.

- PROVEAN Home: <http://provean.jcvi.org/index.php/>
- SIFT Home: <http://sift.jcvi.org/>

Though SIFT is not under active development, the PROVEAN team still provides the SIFT scores in the pre-computed downloads. This package, SIFT.Hsapiens.dbSNP137, contains both SIFT and PROVEAN scores. One notable difference between this and the previous SIFT database package is that keys in SIFT.Hsapiens.dbSNP132 are rs IDs whereas in SIFT.Hsapiens.dbSNP137 they are NCBI dbSNP IDs.

Methods

- Methods : See 'PROVEANdb-class' in the VariantAnnotation package for a complete listing of available methods.
- Creation of Database Tables : This package includes PROVEAN/SIFT predictions for dbSNP build 137 human coding non-synonymous SNPs.
- Source Files :

- Source : <http://provean.jcvi.org/downloads.php>
- Software : PROVEAN 1.1, SIFT 4.0.3
- Databases : PSI-BLAST
- Source Files : `dbsnp137.coding.variants.prediction.tsv.gz` PROVEAN/SIFT predictions for coding snps in dbSNP build 137
- Description : This package contains PROVEAN/SIFT annotations human SNPs included in dbSNP build 137.

Column descriptions

These names are displayed when `columns` is called on the `PROVEANdb` object (i.e., `columns(SIFT.Hsapiens.dbSNP137)`)

- `DBSNPID` : NCBI dbSNP ID
- `VARIANT` : comma separated values of `<chromosome>`, `<position>`, `<reference allele>`, `<variant allele>`, `<comment(optional)>`
- `PROTEINID` : Ensembl protein ID
- `LENGTH` : length of the protein
- `STRAND` : '+' , '-' or NA
- `CODONCHANGE` : codon change including flanking codons
- `POS` : position of amino acid residue affected
- `RESIDUEREf` : reference amino acid residue
- `RESIDUEALT` : variant amino acid residue
- `TYPE` : synonymous | nonsynonymous | frameshift | ...
- `PROVEANSCORE` : PROVEAN score (see http://provean.jcvi.org/about.php#about_1)
- `PROVEANPRED` : deleterious or neutral (cutoff=-2.5)
- `PROVEANNUMSEQ` : number of sequences used for prediction
- `PROVEANNUMCLUST` : number of clusters used for prediction
- `SIFTSCORE` : SIFT score (range 0 to 1)
- `SIFTPRED` : tolerated or damaging (cutoff=0.05)
- `SIFTMEDIAN` : median sequence information used to measure the diversity of the sequences used for prediction
- `SIFTNUMSEQ` : number of sequences used for prediction

Author(s)

Valerie Obenchain <vobencha@fhcrc.org>

References

The PROVEAN tool has replaced SIFT: <http://provean.jcvi.org/about.php>

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the Functional Effect of Amino Acid Substitutions and Indels. PLoS ONE 7(10): e46688.

Choi Y (2012) A Fast Computation of Pairwise Sequence Alignment Scores Between a Protein and a Set of Single-Locus Variants of Another Protein. In Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine (BCB '12). ACM, New York, NY, USA, 414-417.

Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4(7):1073-81

Ng PC, Henikoff S. Predicting the Effects of Amino Acid Substitutions on Protein Function Annu Rev Genomics Hum Genet. 2006;7:61-80.

Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003 Jul 1;31(13):3812-4.

See Also

[PROVEANdb-class](#)

Examples

```
library(SIFT.Hsapiens.dbSNP137)

## metadata
metadata(SIFT.Hsapiens.dbSNP137)

## keys are the DBSNPID (NCBI dbSNP ID)
dbsnp <- keys(SIFT.Hsapiens.dbSNP137)
head(dbsnp)
columns(SIFT.Hsapiens.dbSNP137)

## Return all columns. Note that the key, DBSNPID,
## is always returned.
select(SIFT.Hsapiens.dbSNP137, dbsnp[10])
## subset on keys and cols
cols <- c("VARIANT", "PROVEANPRED", "SIFTPRED")
select(SIFT.Hsapiens.dbSNP137, dbsnp[20:23], cols)
```

Index

* **data**

SIFT.Hsapiens.dbSNP137, [1](#)

* **package**

SIFT.Hsapiens.dbSNP137, [1](#)

PROVEANdb-class, [3](#)

SIFT.Hsapiens.dbSNP137, [1](#)

SIFT.Hsapiens.dbSNP137-package
(SIFT.Hsapiens.dbSNP137), [1](#)