hgu95av2

November 11, 2026

---

hgu95av2 *Bioconductor annotation data package*

Description

The annotation package was built using a downloadable R package - AnnBuilder (download and build your own) from www.bioconductor.org using the following public data sources:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Gene Ontology:\urlhttp://archive.godatabase.org/latest/. Built: 08-Feb-2007

Golden Path:\urlftp://hgdownload.cse.ucsc.edu/goldenPath/currentGenomes/. Built: No build info available.

KEGG:\urlftp://ftp.genome.ad.jp/pub/kegg/tarfiles/pathway.tar.gz. Built: Release 41.1, February 1, 2007

Thu Mar 15 21:19:18 2007

hgu95av2

The function hgu95av2() provides information about the binary data files

---

hgu95av2ACCNUM *Map Manufacturer identifiers to GenBank Accession Numbers*

Description

hgu95av2ACCNUM is an R environment that contains mappings between a manufacturer's identifiers and GenBank accession numbers.

Details

This environment was produced by first mapping the manufacturer's identifiers to Entrez Gene identifiers http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene. The mapped Entrez Gene identifiers then serve as the point of linkage to the GenBank accession numbers.

Each manufacturer identifier maps to a vector containing a GenBank accession number.

Package built: Thu Mar 15 21:19:18 2007

Examples

```
        # Convert to a list
        xx <- as.list(hgu95av2ACCNUM)
        # Remove probe identifiers that do not map to any ACCNUM
        xx <- xx[!is.na(xx)]
        if(length(xx) > 0){
        # Get the ACCNUM for the first five probes
        xx[1:5]
        # Get the first one
        xx[[1]]
}
```

---

hgu95av2CHR *Map Manufacturer IDs to Chromosomes*

Description

hgu95av2CHR is an R environment that provides mappings between a manufacturer identifier and the chromosome that contains the gene of interest.

Details

Each manufacturer identifier maps to a vector of chromosomes. Due to inconsistencies that may exist at the time the environment was built, the vector may contain more than one chromosome (e.g., the identifier may map to more than one chromosome). If the chromosomal location is unknown, the vector will contain an NA.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built: Thu Mar 15 21:19:18 2007

Examples

```
        # Convert to a list
        xx <- as.list(hgu95av2CHR)
        # Remove probe identifiers that do not map to any CHRLOC
        xx <- xx[!is.na(xx)]
        if(length(xx) > 0){
```

```
    # Get the chromosome number for the first five probes
    xx[1:5]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2CHRLENGTHS *A named vector for the length of each of the chromosomes*

Description

hgu95av2CHRLENGTHS provides the length measured in base pairs for each of the chromosomes.

Details

This is a named vector with chromosome numbers as the names and the corresponding lengths for chromosomes as the values.

Total lengths of chromosomes were derived by calculating the number of base pairs on the sequence string for each chromosome.

Annotations were derived base on data privided by:

Golden Path:\urlftp://hgdownload.cse.ucsc.edu/goldenPath/currentGenomes/. Built: No build info available.

Examples

```
tt <- hgu95av2CHRLENGTHS
# Length of chromosome 1
tt["1"]
```

---

hgu95av2CHRLOC *Map Manufacturer IDs to Chromosomal Location*

Description

hgu95av2CHRLOC is an R environment that maps manufacturer identifiers to the starting position of the gene. The position of a gene is measured as the number of base pairs.

Details

Each manufacturer identifier maps to a named vector of chromosomal locations, where the name indicates the chromosome. Due to inconsistencies that may exist at the time the environment was built, these vectors may contain more than one chromosome and/or location. If the chromosomal location is unknown, the vector will contain an NA.

Chromosomal locations on both the sense and antisense strands are measured as the number of base pairs from the p (5' end of the sense strand) to q (3' end of the sense strand) arms. Chromosomal locations on the antisense strand have a leading "-" sign (e. g. -1234567).

Mappings were based on data provided by:

Golden Path:\urlftp://hgdownload.cse.ucsc.edu/goldenPath/currentGenomes/. Built: No build info available.

Package built Thu Mar 15 21:19:18 2007

Examples

```
# Covert to a list
xx <- as.list(hgu95av2CHRLOC)
# Remove probe identifiers that do not map to any CHRLOC
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # Gets the location for the first five probes
    xx[1:5]
    # Gets the first one
    xx[[1]]
}
```

---

hgu95av2ENTREZID *Map between Manufacturer Identifiers and Entrez Gene*

Description

hgu95av2ENTREZID is an R environment that provides mappings between manufacturer identifiers and Entrez Gene identifiers.

Details

Each manufacturer identifier is mapped to a vector of Entrez Gene identifiers. An NA is assigned to those manufacturer identifiers that can not be mapped to an Entrez Gene identifier at this time.

If a given manufacturer identifier can be mapped to different Entrez Gene identifiers from various sources, we attempt to select the common identifiers. If a concensus cannot be determined, we select the smallest identifier.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built: Thu Mar 15 21:19:18 2007

References

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2ENTREZID)
# Remove probe identifiers that do not map to any ENTREZID
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # The ENTREZIDs for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2ENZYME *Map Between Manufacturer IDs and Enzyme Commission (EC) Numbers*

Description

hgu95av2ENZYME is an R environment that provides mappings between manufacturer identifiers and EC numbers.

Details

Each manufacturer identifier maps to a named vector containing the EC number that corresponds to the enzyme produced by that gene. The name corresponds to the manufacturer identifier. If this information is unknown, the vector will contain an NA.

Enzyme Commission numbers are assigned by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology http://www.chem.qmw.ac.uk/iubmb/enzyme/ to allow enzymes to be identified.

An Enzyme Commission number is of the format EC x.y.z.w, where x, y, z, and w are numeric numbers. In hgu95av2ENZYME2PROBE, EC is dropped from the Enzyme Commission numbers.

Enzyme Commission numbers have corresponding names that describe the functions of enzymes in such a way that EC x is a more general description than EC x.y that in turn is a more general description than EC x.y.z. The top level EC numbers and names are listed below:

EC 1 oxidoreductases

EC 2 transferases

EC 3 hydrolases

EC 4 lyases

EC 5 isomerases

EC 6 ligases

The EC name for a given EC number can be viewed at http://www.chem.qmul.ac.uk/iupac/jcbn/index.html#6

Mappings between probe identifiers and enzyme identifiers were obtained using files provided by:

KEGG:\urlftp://ftp.genome.ad.jp/pub/kegg/tarfiles/pathway.tar.gz. Built: Release 41.1, February 1, 2007

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

ftp://ftp.genome.ad.jp/pub/kegg/pathways

Examples

```
        # Convert to a list
        xx <- as.list(hgu95av2ENZYME)
        # Remove probe identifiers that do not map to any enzyme EC number
        xx <- xx[!is.na(xx)]
        if(length(xx) > 0){
        # Gets the enzyme commission numbers for the first five
        #probes
        xx[1:5]
        # Get the first one
        xx[[1]]
}
```

---

hgu95av2ENZYME2PROBE *Map Between Enzyme Commission Numbers and Manufacturer Identifiers*

Description

hgu95av2ENZYME2PROBE is an R environment that maps Enzyme Commission (EC) numbers to manufacturer identifiers.

Details

Each EC number maps to a named vector containing all of the manufacturer identifiers that correspond to the gene that produces that enzyme. The name of the vector corresponds to the EC number.

Enzyme Commission numbers are assigned by the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology http://www.chem.qmw.ac.uk/iubmb/enzyme/ to allow enzymes to be identified.

An Enzyme Commission number is of the format EC x.y.z.w, where x, y, z, and w are numeric numbers. In hgu95av2ENZYME2PROBE, EC is dropped from the Enzyme Commission numbers.

Enzyme Commission numbers have corresponding names that describe the functions of enzymes in such a way that EC x is a more general description than EC x.y that in turn is a more general description than EC x.y.z. The top level EC numbers and names are listed below:

EC 1 oxidoreductases

EC 2 transferases

EC 3 hydrolases

EC 4 lyases

EC 5 isomerases

EC 6 ligases

The EC name for a given EC number can be viewed at http://www.chem.qmul.ac.uk/iupac/jcbn/index.html#6

Mappings were based on data provided by:

KEGG:\urlftp://ftp.genome.ad.jp/pub/kegg/tarfiles/pathway.tar.gz. Built: Release 41.1, February 1, 2007

Package built Thu Mar 15 21:19:18 2007

References

ftp://ftp.genome.ad.jp/pub/kegg/pathways

Examples

```
 # Convert to a list
 xx <- as.list(hgu95av2ENZYME2PROBE)
if(length(xx) > 0){
    # Gets the probe identifiers for the first five enzyme
    #commission numbers
    xx[1:5]
    # Get the first one
    xx[[1]]
 }
```

---

hgu95av2GENENAME *Map Between Manufacturer IDs and Genes*

Description

hgu95av2GENENAME is an R environment that maps manufacturer identifiers to the corresponding gene name.

Details

Each manufacturer identifier maps to a named vector containing the gene name. The vector name corresponds to the manufacturer identifier. If the gene name is unknown, the vector will contain an NA.

Gene names currently include both the official (validated by a nomenclature committee) and preferred names (interim selected for display) for genes. Efforts are being made to differentiate the two by adding a name to the vector.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

Examples

```
# Convert to a list
xx <- as.list(hgu95av2GENENAME)
# Remove probes that do not map to any GENENAME
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # Gets the gene names for the first five probe identifiers
    xx[1:5]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2GO *Map between Manufacturer IDs and Gene Ontology (GO)*

Description

hgu95av2GO is an R environment that provides mappings between manufacturer identifiers and the GO identifiers that they are directly associated with.

Details

Each Entrez Gene identifier is mapped to a list of lists. The names on the outer list are GO identifiers. Each inner list consists of three named elements: GOID, Ontology, and Evidence.

The GOID element matches the GO identifier named in the outer list and is included for convenience when processing the data using 'lapply'.

The Ontology element indicates which of the three Gene Ontology categories this identifier belongs to. The categories are biological process (BP), cellular component (CC), and molecular function (MF).

The Evidence element contains a code indicating what kind of evidence supports the association of the GO identifier to the Entrez Gene id. The evidence codes in use include:

IMP: inferred from mutant phenotype

IGI: inferred from genetic interaction

IPI: inferred from physical interaction

ISS: inferred from sequence similarity

IDA: inferred from direct assay

IEP: inferred from expression pattern

IEA: inferred from electronic annotation

TAS: traceable author statement

NAS: non-traceable author statement

ND: no biological data available

IC: inferred by curator

Mappings between probe identifiers and GO information were obtained through their mappings to Entrez Gene identifiers. NAs are assigned to probe identifiers that can not be mapped to any Gene Ontology information. Mappings between Gene Ontology identifiers an Gene Ontology terms and other information are available in a separate data package named GO.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Gene Ontology:\urlhttp://archive.godatabase.org/latest/. Built: 08-Feb-2007

Package built Thu Mar 15 21:19:18 2007

Examples

```
# Convert to a list
xx <- as.list(hgu95av2GO)
# Remove all the NAs
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
        # Try the firest one
        got <- xx[[1]]
        got[[1]][["GOID"]]
        got[[1]][["Ontology"]]
        got[[1]][["Evidence"]]

}
```

---

hgu95av2GO2ALLPROBES *Map Between Gene Ontology (GO) Identifiers and all Manufacturer Identifiers in the subtree*

Description

hgu95av2GO2ALLPROBES is an R environment that provides mappings between a given GO identifier and all manufactuerer identifiers annotated at that GO term or one of its children in the GO ontology.

Details

GO consists of three ontologies—molecular function (MF), biological process (BP), and cellular component (CC). All ontologies are structured as directed acyclic graphs (DAGs). Each node in each DAG (tree) is a GO term (id) associated with a named vector of manufacturer identifiers. The name associated with each manufacturer id corresponds to the evidence code for that GO identifier. This environment hgu95av2GO2ALLPROBES maps between a given GO identifier and all manufactuerer identifiers annotated at that GO term or one of its children in the GO ontology.

The evidence code indicates what kind of evidence supports the association between the GO and Entrez Gene identifiers. Evidence codes currently in use include:

IMP - inferred from mutant phenotype

IGI - inferred from genetic interaction

IPI - inferred from physical interaction

ISS - inferred from sequence similarity

IDA - inferred from direct assay

IEP - inferred from expression pattern

IEA - inferred from electronic annotation

TAS - traceable author statement

NAS - non-traceable author statement

ND - no biological data available

IC - inferred by curator

A GO identifier may be mapped to the same manufacturer identifier more than once but the evidence code can be different. Mappings between Gene Ontology identifiers and Gene Ontology terms and other information are available in a separate data package named GO.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/

Examples

```
# Convert to a list
xx <- as.list(hgu95av2GO2ALLPROBES)
if(length(xx) > 0){
    # Gets the probe identifiers for the top 2nd and 3nd GO identifiers
    goids <- xx[2:3]
    # Gets all the probe identifiers for the first element of goids
    goids[[1]]
    # Evidence code for the mappings
    names(goids[[1]])
}
```

hgu95av2GO2PROBE *Map Between Gene Ontology (GO) and Manufacturer Identifiers*

Description

hgu95av2GO2PROBE is an R environment that provides mappings between GO identifiers and manufacturer identifiers.

Details

Each GO term maps to a named vector of manufacturer identifiers. The name associated with each manufacturer identifier corresponds to the evidence code for that GO identifier. The evidence code indicates what kind of evidence supports the association between the GO and Entrez Gene identifiers. Evidence codes currently in use include:

IMP - inferred from mutant phenotype

IGI - inferred from genetic interaction

IPI - inferred from physical interaction

ISS - inferred from sequence similarity

IDA - inferred from direct assay

IEP - inferred from expression pattern

IEA - inferred from electronic annotation

TAS - traceable author statement

NAS - non-traceable author statement

ND - no biological data available

IC - inferred by curator

A GO identifier may be mapped to the same probe identifier more than once but the evidence code can be different. Mappings between Gene Ontology identifiers an Gene Ontology terms and other information are available in a separate data package named GO.

Mappings were based on data provided by:

Entrez Gene Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007.

Package built Thu Mar 15 21:19:18 2007

References

ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/

Examples

```
# Convert to a list
xx <- as.list(hgu95av2GO2PROBE)
if(length(xx) > 0){
    # Gets the probe identifiers for the top 2nd and 3nd GO identifiers
    goids <- xx[2:3]
    # Gets the probe identifiers for the first element of goids
    goids[[1]]
    # Evidence code for the mappings
    names(goids[[1]])
}
```

---

hgu95av2MAP *Map between Manufacturer Identifiers and cytogenetic maps/bands*

Description

hgu95av2MAP is an R environment that provides mappings between manufacturer identifiers and cytoband locations.

Details

Each manufacturer identifier is mapped to a vector of cytoband locations. The vector length may be one or longer, if there are multiple reported chromosomal locations for a given gene. An NA is reported for any manufacturer identifiers that cannot be mapped to a cytoband at this time.

Cytogenetic bands for most higher organisms are labeled p1, p2, p3, q1, q2, q3 (p and q are the p and q arms), etc., counting from the centromere out toward the telomeres. At higher resolutions, sub-bands can be seen within the bands. The sub-bands are also numbered from the centromere out toward the telomere. Thus, a label of 7q31.2 indicates that the band is on chromosome 7, q arm, band 3, sub-band 1, and sub-sub-band 2.

The physical location of each band on a chromosome can be obtained from another environment named "organism"CYTOLOC in a separate data package for human(humanCHRLOC), mouse(mouseCHRLOC), and rat(ratCHRLOC).

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.ncbi.nlm.nih.gov

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2MAP)
# Remove probe identifiers that do not map to any cytoband
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # The cytobands for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2OMIM *Map between Manufacturer Identifiers and Mendelian Inheritance in Man (MIM) identifiers*

Description

hgu95av2OMIM is an R environment that provides mappings between manufacturer identifiers and OMIM identifiers.

Details

Each manufacturer identifier is mapped to a vector of OMIM identifiers. The vector length may be one or longer, depending on how many OMIM identifiers the manufacturer identifier maps to. An NA is reported for any manufacturer identifier that cannot be mapped to an OMIM identifier at this time.

OMIM is based upon the book Mendelian Inheritance in Man (V. A. McKusick) and focuses primarily on inherited or heritable genetic diseases. It contains textual information, pictures, and reference information that can be searched using various terms, among which the MIM number is one.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene http://www3.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2OMIM)
# Remove probe identifiers that do not map to any MIM number
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
```

```
    # The MIM numbers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2ORGANISM *The Organism for hgu95av2*

Description

hgu95av2ORGANISM is an R environment that contains a single item: a character string that names the organism for which hgu95av2 was built.

Details

Although the package name is suggestive of the organism for which it was built, hgu95av2ORGANISM provides a simple way to programmatically extract the organism name.

Package built Thu Mar 15 21:19:18 2007

Examples

```
hgu95av2ORGANISM
```

---

hgu95av2PATH *Mappings between probe identifiers and KEGG pathway identifiers*

Description

KEGG (Kyoto Encyclopedia of Genes and Genomes) maintains pathway data for various organisms. hgu95av2PATH maps probe identifiers to the identifiers used by KEGG for pathways in which the genes represented by the probe identifiers are involved

Details

Each KEGG pathway has a name and identifier. Pathway name for a given pathway identifier can be obtained using the KEGG data package that can either be built using AnnBuilder or downloaded from Bioconductor http://www.bioconductor.org.

Graphic presentations of pathways are searchable at urlhttp://www.genome.ad.jp/kegg/pathway.html by using pathway identifiers as keys.

Mappings were based on data provided by:

KEGG:\urlftp://ftp.genome.ad.jp/pub/kegg/tarfiles/pathway.tar.gz. Built: Release 41.1, February 1, 2007

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.genome.ad.jp/kegg/

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2PATH)
# Remove probe identifiers that do not map to any pathway id
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # The pathway identifiers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2PATH2PROBE *Map between Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway identifiers and Manufacturer Identifiers*

Description

hgu95av2PATH2PROBE is an R environment that provides mappings between KEGG identifiers and manufacturer identifiers.

Details

Each KEGG identifier is mapped to a named vector of manufacturer identifiers. The name represents the KEGG identifier and the vector contains all manufacturer identifiers that are found in that particular pathway. An NA is reported for any KEGG identifier that cannot be mapped to any manufacturer identifiers.

Pathway name for a given pathway identifier can be obtained using the KEGG data package that can either be built using AnnBuilder or downloaded from Bioconductor http://www.bioconductor.org.

Graphic presentations of pathways are searchable at http://www.genome.ad.jp/kegg/pathway.html using pathway identifiers as keys.

Mappings were based on data provided by:

KEGG:\urlftp://ftp.genome.ad.jp/pub/kegg/tarfiles/pathway.tar.gz. Built: Release 41.1, February 1, 2007

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.genome.ad.jp/kegg/

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2PATH2PROBE)
# Remove pathway identifiers that do not map to any probe id
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # The probe identifiers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2PFAM *Map Manufacturer IDs to Pfam IDs*

Description

hgu95av2PFAM is an R environment that provides mappings between a manufacturer identifier and the associated Pfam identifiers.

Details

Each manufacturer identifier maps to a named vector of Pfam identifiers. The name for each Pfam identifier is the IPI accession numbe where this Pfam identifier is found.

If the Pfam is a named NA, it means that the associated Entrez Gene id of this manufacturer identifier is found in an IPI entry of the IPI database, but there is no Pfam identifier in the entry.

If the Pfam is a non-named NA, it means that the associated Entrez Gene id of this manufacturer identifier is not found in any IPI entry of the IPI database.

References

Kersey P. J., Duarte J., Williams A., Karavidopoulou Y., Birney E., Apweiler R. The International Protein Index: An integrated database for proteomics experiments. Proteomics 4(7): 1985-1988 (2004) http://www.ebi.ac.uk/IPI/IPIhelp.html

Examples

```
# Convert to a list
xxx <- as.list(hgu95av2PFAM)
# randomly display 10 probes
sample(xxx, 10)
```

hgu95av2PMID *Map between Manufacturer Identifiers and PubMed Identifiers*

Description

hgu95av2PMID is an R environment that provides mappings between manufacturer identifiers and PubMed identifiers.

Details

Each manufacturer identifier is mapped to a named vector of PubMed identifiers. The name associated with each vector corresponds to the manufacturer identifier. The length of the vector may be one or greater, depending on how many PubMed identifiers a given manufacturer identifier is mapped to. An NA is reported for any manufacturer identifier that cannot be mapped to a PubMed identifier.

Titles, abstracts, and possibly full texts of articles can be obtained from PubMed by providing a valid PubMed identifier. The pubmed function of annotate can also be used for the same purpose.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2PMID)
if(length(xx) > 0){
    # The probe identifiers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
    if(interactive() && !is.null(xx[[1]]) && !is.na(xx[[1]])
       && require(annotate)){
        # Gets article information as XML files
        xmls <- pubmed(xx[[1]], disp = "data")
        # Views article information using a browser
        pubmed(xx[[1]], disp = "browser")
    }
}
```

hgu95av2PMID2PROBE *Map between PubMed Identifiers and Manufacturer Identifiers*

Description

hgu95av2PMID2PROBE is an R environment that provides mappings between PubMed identifiers and manufacturer identifiers.

Details

Each PubMed identifier is mapped to a named vector of manufacturer identifiers. The name represents the PubMed identifier and the vector contains all manufacturer identifiers that are represented by that PubMed identifier. The length of the vector may be one or longer, depending on how many manufacturer identifiers are mapped to a given PubMed identifier.

Titles, abstracts, and possibly full texts of articles can be obtained from PubMed by providing a valid PubMed identifier. The pubmed function of annotate can also be used for the same purpose

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2PMID2PROBE)
if(length(xx) > 0){
    # The probe identifiers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
    if(interactive() && require(annotate)){
        # Gets article information as XML files for a PubMed id
        xmls <- pubmed(names(xx)[1], disp = "data")
        # Views article information using a browser
        pubmed(names(xx)[1], disp = "browser")
    }
}
```

hgu95av2PROSITE *Map Manufacturer IDs to PROSITE ID*

Description

hgu95av2PROSITE is an R environment that provides mappings between a manufacturer identifier and the associated PROSITE identifiers.

Details

Each manufacturer identifier maps to a named vector of PROSITE identifiers. The name for each PROSITE identifier is the IPI accession numbe where this PROSITE identifier is found.

If the PROSITE is a named NA, it means that the associated Entrez Gene id of this manufacturer identifier is found in an IPI entry of the IPI database, but there is no PROSITE identifier in the entry.

If the PROSITE is a non-named NA, it means that the associated Entrez Gene id of this manufacturer identifier is not found in any IPI entry of the IPI database.

References

Kersey P. J., Duarte J., Williams A., Karavidopoulou Y., Birney E., Apweiler R. The International Protein Index: An integrated database for proteomics experiments. Proteomics 4(7): 1985-1988 (2004) http://www.ebi.ac.uk/IPI/IPIhelp.html

Examples

```
# Convert to a list
xxx <- as.list(hgu95av2PROSITE)
# randomly display 10 probes
xxx[sample(1:length(xxx), 10)]
```

---

hgu95av2QC *Quality control information for hgu95av2*

Description

hgu95av2QC is an R environment that provides quality control information for hgu95av2

Details

This file contains quality control information that can be displayed by typing hgu95av2() after loading the package using library(hgu95av2). The follow items are included:

Date built: Date when the package was built.

Number of probes: total number of probes included

Probe number missmatch: if the total number of probes of any of the data file is different from a base file used to check the data files the name of the data file will be listed

Probe missmatch: if any of probes in a data file missmatched that of the base file, the name of the data file will be listed

Mappings found for probe based files: number of mappings obtained for the total number of probes

Mappings found for non-probe based files: total number of mappings obtained

hgu95av2QCDATA *Quality control information for hgu95av2*

Description

hgu95av2QC is a R list that provides quality control information for hgu95av2

Details

This file contains quality control information that can be displayed by typing hgu95av2() after loading the package using library(hgu95av2). The follow items are included:

Date built: Date when the package was built.

Session infromation

Number of probes: total number of probes included

Probe number missmatch: if the total number of probes of any of the data file is different from a base file used to check the data files the name of the data file will be listed

Probe missmatch: if any of probes in a data file missmatched that of the base file, the name of the data file will be listed

Mappings found for probe based files: number of mappings obtained for the total number of probes

Mappings found for non-probe based files: total number of mappings obtained

hgu95av2REFSEQ *Map between Manufacturer Identifiers and RefSeq Identifiers*

Description

hgu95av2REFSEQ is an R environment that provides mappings between manufacturer identifiers and RefSeq identifiers.

Details

Each manufacturer identifier is mapped to a named vector of RefSeq identifiers. The name represents the manufacturer identifier and the vector contains all RefSeq identifiers that can be mapped to that manufacturer identifier. The length of the vector may be one or greater, depending on how many RefSeq identifiers a given manufacturer identifier can be mapped to. An NA is reported for any manufacturer identifier that cannot be mapped to a RefSeq identifier at this time.

RefSeq identifiers differ in format according to the type of record the identifiers are for as shown below:

NG_XXXXX: RefSeq accessions for genomic region (nucleotide) records

NM_XXXXX: RefSeq accessions for mRNA records

NC_XXXXX: RefSeq accessions for chromosome records

NP_XXXXX: RefSeq accessions for protein records

XR_XXXXX: RefSeq accessions for model RNAs that are not associated with protein products

XM_XXXXX: RefSeq accessions for model mRNA records

XP_XXXXX: RefSeq accessions for model protein records

Where XXXXX is a sequence of integers.

NCBI http://www.ncbi.nlm.nih.gov/RefSeq/ allows users to query the RefSeq database using RefSeq identifiers.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.ncbi.nlm.nih.gov http://www.ncbi.nlm.nih.gov/RefSeq/

Examples

```
    # Convert the environment to a list
    xx <- as.list(hgu95av2REFSEQ)
    # Remove probe identifiers that do not map to any RefSeq
    xx <- xx[!is.na(xx)]
    if(length(xx) > 0){
    # The RefSeq for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2SYMBOL *Map between Manufacturer Identifiers and Gene Symbols*

Description

hgu95av2SYMBOL is an R environment that provides mappings between manufacturer identifiers and gene abbreviations.

Details

Each manufacturer identifier is mapped to an abbreviation for the corresponding gene. An NA is reported if there is no known abbreviation for a given gene.

Symbols typically consist of 3 letters that define either a single gene (ABC) or multiple genes (ABC1, ABC2, ABC3). Gene symbols can be used as key words to query public databases such as Entrez Gene.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2SYMBOL)
if(length(xx) > 0){
    # The symbols for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

hgu95av2UNIGENE *Map between Manufacturer Identifiers and UniGene cluster identifiers*

Description

hgu95av2UNIGENE is an R environment that provides mappings between manufacturer identifiers and UniGene identifiers.

Details

Each manufacturer identifier is mapped to a UniGene identifier. An NA is reported if the manufacturer identifier cannot be mapped to UniGene at this time.

A UniGene identifier represents a cluster of sequences of a gene. Using UniGene identifiers one can query the UniGene database for information about the sequences or the Entrez Gene database for information about the genes.

Mappings were based on data provided by:

Entrez Gene:\urlftp://ftp.ncbi.nlm.nih.gov/gene/DATA/. Built: Source data downloaded from Entrez Gene on Thu Mar 15 20:55:24 2007

Package built Thu Mar 15 21:19:18 2007

References

http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene

Examples

```
# Convert the environment to a list
xx <- as.list(hgu95av2UNIGENE)
# Remove probe identifiers that do no map to any UniGene id
xx <- xx[!is.null(xx)]
if(length(xx) > 0){
    # The UniGene identifiers for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

Index

∗ **datasets**
  hgu95av2, 1
  hgu95av2ACCNUM, 1
  hgu95av2CHR, 2
  hgu95av2CHRLENGTHS, 3
  hgu95av2CHRLOC, 3
  hgu95av2ENTREZID, 4
  hgu95av2ENZYME, 5
  hgu95av2ENZYME2PROBE, 6
  hgu95av2GENENAME, 7
  hgu95av2GO, 8
  hgu95av2GO2ALLPROBES, 9
  hgu95av2GO2PROBE, 11
  hgu95av2MAP, 12
  hgu95av2OMIM, 13
  hgu95av2ORGANISM, 14
  hgu95av2PATH, 14
  hgu95av2PATH2PROBE, 15
  hgu95av2PFAM, 16
  hgu95av2PMID, 17
  hgu95av2PMID2PROBE, 18
  hgu95av2PROSITE, 18
  hgu95av2QC, 19
  hgu95av2QCDATA, 20
  hgu95av2REFSEQ, 20
  hgu95av2SYMBOL, 21
  hgu95av2UNIGENE, 22

hgu95av2, 1
hgu95av2ACCNUM, 1
hgu95av2CHR, 2
hgu95av2CHRLENGTHS, 3
hgu95av2CHRLOC, 3
hgu95av2ENTREZID, 4
hgu95av2ENZYME, 5
hgu95av2ENZYME2PROBE, 6
hgu95av2GENENAME, 7
hgu95av2GO, 8
hgu95av2GO2ALLPROBES, 9
hgu95av2GO2PROBE, 11
hgu95av2LOCUSID *(*hgu95av2ENTREZID*)*, 4
hgu95av2MAP, 12
hgu95av2MAPCOUNTS *(*hgu95av2QC*)*, 19
hgu95av2OMIM, 13
hgu95av2ORGANISM, 14
hgu95av2PATH, 14
hgu95av2PATH2PROBE, 15
hgu95av2PFAM, 16
hgu95av2PMID, 17
hgu95av2PMID2PROBE, 18
hgu95av2PROSITE, 18
hgu95av2QC, 19
hgu95av2QCDATA, 20
hgu95av2REFSEQ, 20
hgu95av2SYMBOL, 21
hgu95av2UNIGENE, 22