

hgu2beta7

March 19, 2011

hgu2beta7

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:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Gene Ontology:http://www.godatabase.org/dev/database/archive/daily/go_daily-termdb.rdf-xml.gz. Built: http://www.godatabase.org/dev/database/archive/daily/go_daily-termdb.rdf-xml.gz

Golden Path:<http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/>. Built: hg17

KEGG:<ftp://ftp.genome.ad.jp/pub/kegg/pathways>. Built: Release 33.0 (January 2005)

Wed Feb 16 23:32:20 2005

hgu2beta7

The function `hgu2beta7()` provides information about the binary data files

hgu2beta7ACCNUM

Mappings between probe identifiers and manufacturer or user provided identifiers

Description

hgu2beta7ACCNUM contains mappings between probe ids to identifiers provided by a manufacturer or user at the time when hgu2beta7 was built. The manufacture or user provided ids were used to link probe ids to annotation data available from various public data sources to obtain the other annotations contained in hgu2beta7.

Details

In order to build an annotated data package, a data file with a column for probe ids and another for the matching manufacturer/user provided ids is required. This data file is used as the base for mapping probe ids to LocusLink identifiers (unique ids used by NCBI <http://www.ncbi.nlm.nih.gov/LocusLink/> to represent genetic loci and link to curated sequence and descriptive information) through the manufacturer/user provided ids. The mapped LocusLink ids then serve as the point of linkage to other annotation data provided by various public data sources.

hgu2beta7ACCNUM contains the mappings of the base file and the derived mappings between probe ids to other annotation data are contained by other annotation data files in hgu2beta7.

Valid manufacturer/user provided ids currently include GenBank accession numbers (identifiers obtained after submitting a sequence to GenBank, EMBL or DDBJ), UniGene ids (identifiers used by NCBI to represent clusters of sequences of a unique gene), RefSeq ids (identifiers for non-redundant sequences curated by NCBI), or IMAGE clone ids (identifiers given to clones from shared arrayed cDNA libraries by The Integrated Molecular Analysis of Gene Expression (IMAGE) Consortium).

For all the Affymetrix chips, the manufacturer/user provided ids are GenBank accession numbers.

Package built: Wed Feb 16 23:32:20 2005

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7ACCNUM)
# Remove probe ids 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]]
}
```

hgu2beta7CHR

Mappings between probe identifiers and the chromosome numbers the genes represented by the probe ids reside

Description

Each chromosome of a given organism is assigned a number or letter for labeling purpose. hgu2beta7CHR maps probe ids to the numbers or letters labeling chromosomes where the genes corresponding to the probe ids reside

Details

Since a given gene may be found in more than one chromosomes, each probe id is mapped to a character vector of length 1 or greater. Probe ids that can not be mapped to any chromosome at the time when the package was built are assigned NAs.

Mappings were based on data provided by:

LocusLink:ftp://ftp.ncbi.nlm.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built: Wed Feb 16 23:32:20 2005

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7CHR)
# Remove probe ids 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]]
}
```

hgu2beta7CHRENGTHS

A named vector for the length of each of the chromosomes

Description

hgu2beta7CHRENGTHS 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 provided by:

Golden Path:<http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/>. Built: hg17

Examples

```
tt <- hgu2beta7CHRENGTHS
# Length of chromosome 1
tt["1"]
```

hgu2beta7CHRLOC

Mappings between probe identifiers and the chromosomal locations of the genes represented by the probe identifiers

Description

hgu2beta7CHRLOC maps probe ids to the starting positions of the genes represented by the probe ids on chromosomes. The position of a gene is measured as the number of base pairs.

Details

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). NAs are assigned to probe ids whose chromosomal locations are not known.

Since a given gene may have more than one chromosomal locations, each probe id is mapped to a named vector of length 1 or greater. The names of each vector give the chromosome number where the genes reside. When a gene can be mapped to a given chromosome but the exact location of the gene can not be determined with confidence, "random" is appended to the end of the name for a chromosomal location value.

Mappings were based on data provided by:

Golden Path:<http://hgdownload.cse.ucsc.edu/goldenPath/hg17/database/>. Built: hg17

Package built Wed Feb 16 23:32:20 2005

Examples

```
# Covert to a list
xx <- as.list(hgu2beta7CHRLOC)
# Remove probe ids 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]]
}
```

hgu2beta7ENZYME

Mappings between probe identifiers and the enzyme commission numbers, to which the products of the genes represented by the probe ids correspond

Description

Genes code for proteins/enzymes that take part in various biological processes. hgu2beta7ENZYME maps probe ids to Enzyme Commission numbers for the enzymes the genes represented by the probe ids code for. 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.

Details

Each probe id is mapped to one or more Enzyme Commission numbers. An Enzyme Commission number is of the format EC x.y.z.w, where x, y, z, and w are numeric numbers. In hgu2beta7ENZYME, 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

hgu2beta7ENZYME only provides mappings to EC numbers. The EC name for a given EC number can be viewed at <http://www.chem.qmul.ac.uk/iupac/jcfn/index.html#6>

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

KEGG:<ftp://ftp.genome.ad.jp/pub/kegg/pathways>. Built: Release 33.0 (January 2005)

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

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

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7ENZYME)
# Remove probe ids 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]]
}
```

hgu2beta7ENZYME2PROBE

Mappings between Enzyme Commission numbers and probe identifiers

Description

hgu2beta7ENZYME2PROBE maps Enzyme Commission numbers to probe identifiers representing genes whose products exhibit the functions of the enzymes labeled by the Commission Numbers

Details

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 hgu2beta7ENZYME2PROBE, 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/jcbon/index.html#6>

Mappings were based on data provided by:

KEGG:<ftp://ftp.genome.ad.jp/pub/kegg/pathways>. Built: Release 33.0 (January 2005)

Package built Wed Feb 16 23:32:20 2005

References

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

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7ENZYME2PROBE)
if(length(xx) > 0){
  # Gets the probe ids for the first five enzyme
  #commission numbers
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```

hgu2beta7GENENAME *Mappings between probe identifiers and the names of genes the probe ids correspond to*

Description

hgu2beta7GENENAME maps probe identifiers to the names of genes corresponding to the probe ids. Gene names are the descriptions used for gene reports that have been validated by a nomenclature committee or interim selected for display

Details

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:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7GENENAME)
# 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 ids
  xx[1:5]
  # Get the first one
  xx[[1]]
}
```

hgu2beta7GO

Mappings between probe identifiers and Gene Ontology information

Description

The Gene Ontology Consortium maintains a controlled vocabulary of defined terms to describe gene products. hgu2beta7GO maps probe identifiers to Gene Ontology information including the ids, evidence codes, and ontology categories of the terms for the genes represented by the probe ids.

Details

Each probe id is mapped to a list containing one (mapped to none or one set of GO information) or more (mapped to more than one sets of GO information) elements. When a probe id is mapped to at least one set of GO information, each element of the list contains a sub list of three elements named "GOID", "Ontology", and "Evidence". The value for element "GOID" gives the Gene Ontology identifier the key probe id corresponds to. The value for element "Ontology" can be an abbreviation of MF (molecular function), BP (biological process), or CC (cellular component) for the GO ontology category the GO id belongs to. The value for element "Evidence" contains an evidence code indicating what kind of evidence is found to support the association of the GO id to the gene represented by the key probe 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 ids and GO information were obtained through their mappings to LocusLink ids. NAs are assigned to probe identifiers that can not be mapped to any Gene Ontology information. Mappings between Gene Ontology ids and Gene Ontology terms and other information are available in a separate data package named GO.

Mappings were based on data provided by:

LocusLink: ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Gene Ontology: http://www.godatabase.org/dev/database/archive/daily/go_daily-termdb.rdf-xml.gz. Built: http://www.godatabase.org/dev/database/archive/daily/go_daily-termdb.rdf-xml.gz

Package built Wed Feb 16 23:32:20 2005

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7GO)
# Remove all the NAs
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # Try the first one
  got <- xx[[1]]
  got[[1]][["GOID"]]
  got[[1]][["Ontology"]]
  got[[1]][["Evidence"]]
}
```

hgu2beta7GO2ALLPROBES

Mappings between GO identifiers and all the probe identifiers representing genes associated with the GO ids

Description

The Gene Ontology (GO) Consortium maintains a controlled vocabulary of defined terms to describe gene products. Each of the terms is assigned a GO id that can be associated with a group of genes directly or indirectly through the children GO ids of the GO id. hgu2beta7GO2ALLPROBES maps GO ids to genes (represented by probe identifiers) that are directly or indirectly mapped to the GO ids by traversing the GO tree.

Details

GO terminologies are presented as a directed acyclic graph with root nodes having more general terms than their children nodes. Therefore, a GO id can be mapped to genes corresponding to the term of that particular GO id as well as genes corresponding to the terms of all the children GO ids. hgu2beta7GO2ALLPROBES maps a GO id to all the probe ids representing genes mapped by traversing the GO tree. The mapped probe ids are named vectors. Names of the vectors are the evidence codes indicating what kind of evidence is found to support the association between GO ids and genes.

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

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

Mappings were based on data provided:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7GO2ALLPROBES)
if(length(xx) > 0){
  # Gets the probe ids for the top 2nd and 3rd GO ids
  goids <- xx[2:3]
  # Gets all the probe ids for the first element of goids
  goids[[1]]
  # Evidence code for the mappings
  names(goids[[1]])
}
```

hgu2beta7GO2PROBE *Mappings between GO identifiers and probe identifiers representing genes that are directly associated with the GO ids*

Description

The Gene Ontology (GO) Consortium maintains a controlled vocabulary of defined terms to describe gene products. Each of the terms is assigned a GO id that can be associated with a group of genes. hgu2beta7GO2PROBE maps each GO id to the probe identifiers representing genes that are associated with the GO ids

Details

GO terminologies are presented as a directed acyclic graph with root nodes having more general terms than their children nodes. hgu2beta7GO2PROBE only maps GO ids to probe ids associated with nodes specified by the GO ids but not their children nodes. The mapped probe ids are named vectors. Names of the vectors are the evidence codes indicating what kind of evidence is found to support the association between GO ids and genes.

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

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

Mappings were based on data provided by:

LocusLink LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz.
Built: February 16, 2005.

Package built Wed Feb 16 23:32:20 2005

References

ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz

Examples

```
# Convert to a list
xx <- as.list(hgu2beta7GO2PROBE)
if(length(xx) > 0){
  # Gets the probe ids for the top 2nd and 3rd GO ids
  goids <- xx[2:3]
  # Gets the probe ids for the first element of goids
  goids[[1]]
  # Evidence code for the mappings
  names(goids[[1]])
}
```

hgu2beta7GRIF

Mappings between probe identifiers and the PubMed identifiers for the publication where functional annotation of genes corresponding to the probe ids were made

Description

NCBI <http://www.ncbi.nlm.nih.gov/projects/GeneRIF/GeneRIFhelp.html> allows users to add functional annotation of genes by providing a PubMed identifier for the publication where the annotation was initially made and a concise phrase describing the function. hgu2beta7GRIF maps probe identifiers to the PubMed identifiers for the publication reporting the functions of genes corresponding to the probe identifiers

Details

Probe ids are only mapped to the PubMed ids. Efforts are being made to also include the concise phrase describing the function of a gene.

Mappings were based on data provided by:

LocusLink: ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built: Wed Feb 16 23:32:20 2005

References

<http://www.ncbi.nlm.nih.gov/LocusLink>

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7GRIF)
# Remove probe ids that do not map to any PubMed id
xx <- xx[!is.na(xx)]
if (length(xx) > 0) {
  # The unique PubMed ids for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}
```

hgu2beta7LOCUSID

Mappings between probe identifiers and LocusLink identifiers

Description

LocusLink identifiers are unique ids used by NCBI <http://www.ncbi.nlm.nih.gov/LocusLink/> to represent genetic loci and link to curated sequence and descriptive information. hgu2beta7LOCUSID maps probe identifiers to LocusLink identifiers for genes represented by the probe identifiers.

Details

Mappings of probe ids to LocusLink ids are achieved through the linkage between probe ids to manufacturer/user provided ids contained in hgu2beta7ACCNUM. LocusLink ids are first mapped to manufacturer/user provided ids using file(s) provided by Locuslink and in some cases UniGene and other available sources and then linked to probe ids based on the mappings between probe ids and manufacturer/user provided ids.

When a given probe id was mapped to different LocusLink ids by different sources, the one agreed by most sources was selected. When no agreement was achieved among sources, the one of the smallest numeric value was selected.

Mappings were based on data provided by:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built: Wed Feb 16 23:32:20 2005

References

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

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7LOCUSID)
# Remove probe ids that do not map to any LOCUSID
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # The LOCUSIDs for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}
```

hgu2beta7MAP

Mappings between probe identifiers and cytogenetic maps/bands

Description

When viewed using a microscope and special stains a chromosome is divided into regions, or cytogenetic bands, of transverse alternating light and dark or fluorescent and nonfluorescent bands. hgu2beta7MAP maps probe identifiers to the labels of cytogenetic bands within chromosomes where genes represented by the probe ids are located

Details

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.

A given probe id may be mapped to one or more cytogenetic bands depending on whether genes represented by probe ids have only one or more chromosomal locations. Different genes may also be mapped to the same cytogenetic bands.

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:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

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

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7MAP)
# Remove probe ids 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]]
}
```

hgu2beta7OMIM

Mappings between probe identifiers and Mendelian Inheritance in Man (MIM) identifiers

Description

The OMIM database (Online Mendelian Inheritance in Man) <http://www3.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM> contains data for human genes and genetic disorders. hgu2beta7OMIM maps probe identifiers to MIM numbers that can be used to search the database

Details

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:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

<http://www.ncbi.nlm.nih.gov/LocusLink>

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7OMIM)
# Remove probe ids 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]]
}
```

hgu2beta7ORGANISM *A character string for the name of the organism*

Description

This is a character string for the name of the organism of concern. The name of the organism can be viewed by typing the name of the binary file that will be loaded when the package is loaded by calling `library(hgu2beta7)`

Details

Although the name of the package is often suggestive of the organism the package concerns, having an environment object for organism name make it easier to get the name of the organism programmatically.

Package built Wed Feb 16 23:32:20 2005

Examples

```
hgu2beta7ORGANISM
```

hgu2beta7PATH *Mappings between probe identifiers and KEGG pathway identifiers*

Description

KEGG (Kyoto Encyclopedia of Genes and Genomes) maintains pathway data for various organisms. hgu2beta7PATH 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 id. Pathway name for a given pathway id 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> by using pathway ids as keys.

Mappings were based on data provided by:

KEGG:<ftp://ftp.genome.ad.jp/pub/kegg/pathways>. Built: Release 33.0 (January 2005)

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

References

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

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7PATH)
# Remove probe ids that do not map to any pathway id
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # The pathway ids for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}
```

hgu2beta7PATH2PROBE

Mappings between KEGG pathway identifiers and probe identifiers

Description

KEGG (Kyoto Encyclopedia of Genes and Genomes) maintains pathway data for various organisms. Each pathway has a name, pathway id, and graphical presentation of the pathway. hgu2beta7PATH2PROBE maps KEGG pathway ids to probe identifiers representing genes whose products participate in the interactions of the pathway

Details

Pathway name for a given pathway id 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 ids as keys.

Mappings were based on data provided by:

KEGG:<ftp://ftp.genome.ad.jp/pub/kegg/pathways>. Built: Release 33.0 (January 2005)

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

References

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

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7PATH2PROBE)
# Remove pathway ids that do not map to any probe id
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
  # The probe ids for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}
```

hgu2beta7PMID

*Mappings between probe identifiers and PubMed identifiers***Description**

PubMed <http://www.ncbi.nlm.nih.gov/PubMed/> includes over 14 million citations for biomedical articles back to the 1950's. Each article has a unique PubMed identifier. hgu2beta7PMID maps probe identifiers to PubMed identifier for publications that cited the genes represented by the probe identifiers

Details

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

Mappings were based on data provided by:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

<http://www.ncbi.nlm.nih.gov/LocusLink>

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7PMID)
if(length(xx) > 0){
  # The probe ids 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")
  }
}
```

hgu2beta7PMID2PROBE

Mappings between PubMed identifiers and probe identifiers

Description

PubMed <http://www.ncbi.nlm.nih.gov/PubMed/> includes over 14 million citations for biomedical articles back to the 1950's. Each article has a unique PubMed identifier. hgu2beta7PMID2PROBE maps PubMed identifiers to probe identifiers representing genes cited by the publications identified by the PubMed identifiers

Details

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

Mappings were based on data provided by:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7PMID2PROBE)
if(length(xx) > 0){
  # The probe ids 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")
  }
}
```

hgu2beta7QC

Quality control information for hgu2beta7

Description

A data file containing statistics for all the data files in this package. The data are used for quality control purpose

Details

This file contains quality control information that can be displayed by typing `hgu2beta7()` after loading the package using `library(hgu2beta7)`. The follow items are included:

Date built: Date when the package was built.

Number of probes: total number of probes included

Probe number mismatch: 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 mismatch: if any of probes in a data file mismatched 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

hgu2beta7REFSEQ *Mappings between probe identifiers and RefSeq identifiers*

Description

The Reference Sequence (RefSeq) database contains curated non-redundant set of sequences for genomic DNA, RNA, and protein for various organisms. `hgu2beta7REFSEQ` maps probe identifiers to all the RefSeq identifiers that are mapped by NCBI to genes represented by the probe identifiers

Details

RefSeq ids differ in format according to the type of record the ids 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 ids.

Mappings were based on data provided by:

LocusLink: ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

NCBI <http://www.ncbi.nlm.nih.gov> and RefSeq <http://www.ncbi.nlm.nih.gov/RefSeq/>

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7REFSEQ)
# Remove probe ids 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]]
}
```

hgu2beta7SUMFUNC *Mappings between probe identifiers and summaries of the functions of genes represented by the probe identifiers*

Description

hgu2beta7SUMFUNC maps probe identifiers to brief summaries of functions of the products of genes represented by the probe identifiers

Details

Annotation of genes by brief summaries of functions of the products of genes represented by the probe identifiers were supported by earlier builds of the LocusLink source data but not so lately. This annotation may have to be dropped in the future.

Mappings were based on data provided by:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

<http://www.ncbi.nlm.nih.gov/LocusLink>

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7SUMFUNC)
if(length(xx) > 0){
# The function summaries for the first two elements of XX
xx[1:2]
# Get the first one
xx[[1]]
}
```

hgu2beta7SYMBOL *Mappings between probe identifiers and gene symbols*

Description

When new genes are reported, gene symbols that are (hopefully) validated by an appropriate nomenclature committee are used to label the genes. hgu2beta7SYMBOL maps probe identifiers to the symbols used to report genes represented by the probe identifiers

Details

hgu2beta7SYMBOL maps probe ids to either the official (validated by a nomenclature committee) and preferred names (interim selected for display). Efforts are being made to differentiate the two.

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 LocusLink.

Mappings were based on data provided by:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

<http://www.ncbi.nlm.nih.gov/LocusLink>

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7SYMBOL)
if(length(xx) > 0){
  # The symbols for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}
```

hgu2beta7UNIGENE *Mappings between probe identifiers and UniGene cluster identifiers*

Description

NCBI partitions sequences submitted to GenBank into non-redundant sets of clusters. Each cluster contains sequences that represent a unique gene and is assigned a unique UniGene id. hgu2beta7UNIGENE maps probe identifiers to UniGene cluster identifiers for genes represented by the probe ids

Details

A UniGene id represents a cluster of sequences of a gene. Using UniGene ids one can query the UniGene database for information about the sequences or the LocusLink database for information about the genes.

Mappings were based on data provided by:

LocusLink:ftp://ftp.ncbi.nih.gov/refseq/LocusLink/LL_tmpl.gz. Built: February 16, 2005

Package built Wed Feb 16 23:32:20 2005

References

<http://www.ncbi.nlm.nih.gov/LocusLink>

Examples

```
# Convert the environment to a list
xx <- as.list(hgu2beta7UNIGENE)
# Remove probe ids that do no map to any UniGene id
xx <- xx[!is.null(xx)]
if(length(xx) > 0){
  # The UniGene ids for the first two elements of XX
  xx[1:2]
  # Get the first one
  xx[[1]]
}
```

Index

*Topic datasets

- hgu2beta7, [1](#)
 - hgu2beta7ACCNUM, [1](#)
 - hgu2beta7CHR, [2](#)
 - hgu2beta7CHRLNGTHS, [3](#)
 - hgu2beta7CHRLOC, [3](#)
 - hgu2beta7ENZYME, [4](#)
 - hgu2beta7ENZYME2PROBE, [5](#)
 - hgu2beta7GENENAME, [6](#)
 - hgu2beta7GO, [7](#)
 - hgu2beta7GO2ALLPROBES, [8](#)
 - hgu2beta7GO2PROBE, [9](#)
 - hgu2beta7GRIF, [11](#)
 - hgu2beta7LOCUSID, [11](#)
 - hgu2beta7MAP, [12](#)
 - hgu2beta7OMIM, [13](#)
 - hgu2beta7ORGANISM, [14](#)
 - hgu2beta7PATH, [14](#)
 - hgu2beta7PATH2PROBE, [15](#)
 - hgu2beta7PMID, [16](#)
 - hgu2beta7PMID2PROBE, [17](#)
 - hgu2beta7QC, [17](#)
 - hgu2beta7REFSEQ, [18](#)
 - hgu2beta7SUMFUNC, [19](#)
 - hgu2beta7SYMBOL, [20](#)
 - hgu2beta7UNIGENE, [20](#)
-
- hgu2beta7, [1](#)
 - hgu2beta7ACCNUM, [1](#)
 - hgu2beta7CHR, [2](#)
 - hgu2beta7CHRLNGTHS, [3](#)
 - hgu2beta7CHRLOC, [3](#)
 - hgu2beta7ENZYME, [4](#)
 - hgu2beta7ENZYME2PROBE, [5](#)
 - hgu2beta7GENENAME, [6](#)
 - hgu2beta7GO, [7](#)
 - hgu2beta7GO2ALLPROBES, [8](#)
 - hgu2beta7GO2PROBE, [9](#)
 - hgu2beta7GRIF, [11](#)
 - hgu2beta7LOCUSID, [11](#)
 - hgu2beta7MAP, [12](#)
 - hgu2beta7MAPCOUNTS (*hgu2beta7QC*), [17](#)
 - hgu2beta7OMIM, [13](#)