gahgu95e

April 7, 2009

R topics documented:

---

gahgu95eALIASES   *Map between custom probeset Identifiers and alternative Gene Symbols*

Description

gahgu95eALIASES is an R environment that provides mappings between custom probeset identifiers and alternative Gene Symbols.

Details

Each identifier is mapped to a named vector containing alternative Gene Symbols for the corresponding gene. An NA is reported if there is no alternative Gene Symbols for a given gene.

Each element of aliases vectors is named with a reference to databases reporting the corresponding alias. The databases names are included in square brakets and could be one or more of the following ones: HGNC, UniProt, SwissProt, TrEMBL, NCBI, GDB, OMIM, GeneLoc

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:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eALIASES)
if(length(xx) > 0){
    # The aliases symbols for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eCHR   *Map custom probeset IDs to Chromosomes*

Description

gahgu95eCHR is an R environment that provides mappings between a custom probeset identifier and the chromosome that contains the gene of interest.

Details

Each identifier maps to a chromosome. If the chromosomal location is unknown, the vector will contain an NA.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built: Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert to a list
xx <- as.list(gahgu95eCHR)
# Remove probe identifiers that do not map to any chromosome
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]]
    }
```

```
gahgu95eDESCRIPTIONS
```
Map between custom probeset Identifiers and gene descriptions

Description

gahgu95eDESCRIPTIONS is an R environment that provides mappings between custom probeset identifiers and corresponding gene description.

Details

Each identifier is mapped to a named vector containing gene description (also called "gene title" or "gene name") for the corresponding gene. An NA is reported if there is no gene description for a given gene.

Each element of named vector contains alternative description for the corresponding gene and is named with a reference to databases reporting the corresponding description. The databases names are included in square brakets and could be one or more of the following ones: HGNC, UniProt, SwissProt, TrEMBL, NCBI, GDB, OMIM, GeneLoc

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eDESCRIPTIONS)
if(length(xx) > 0){
    # The descriptions for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

gahgu95eENTREZID   *Map between custom probeset Identifiers and Entrez Gene*

Description

gahgu95eENTREZID 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.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built: Mon Aug 25 17:16:58 2008

References

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

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eENTREZID)
# 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]]
}
```

gahgu95eGENECARDSID

Map between custom probeset Identifiers and Genecards Identifiers

Description

gahgu95eGENECARDSID is an R environment that provides mappings between custom probeset identifiers and Genecards identifiers.

Details

Each identifier is mapped to a of Genecards identifier.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built: Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eGENECARDSID)
# Remove probe identifiers that do not map to any GENECARDSID
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # The GENECARDSIDs for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eGENEEND	*Map custom probeset IDs to gene end position*

Description

gahgu95eGENEEND is an R environment that provides mappings between a custom probeset identifier and the base pair ending position of gene onto corresponding chromosome.

Details

Each identifier maps to a gene end position. If the gene end position is unknown, the vector will contain an `NA`.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built: Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert to a list
xx <- as.list(gahgu95eGENEEND)
# Remove probe identifiers that do not map to any gene end position
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # Get the base pair number for gene end position for the first five probes
    xx[1:5]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eGENESTART   *Map custom probeset IDs to gene start position*

Description

gahgu95eGENESTART is an R environment that provides mappings between a custom probeset identifier and the base pair starting position of gene onto corresponding chromosome.

Details

Each identifier maps to a gene start position. If the gene start position is unknown, the vector will contain an NA.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built: Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert to a list
xx <- as.list(gahgu95eGENESTART)
# Remove probe identifiers that do not map to any gene start position
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # Get the base pair number for gene start position for the first five probes
    xx[1:5]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eGO   *Map between custom probeset IDs and Gene Ontology (GO)*

Description

gahgu95eGO is an R environment that provides mappings between custom probeset identifiers and the GO identifiers that they are directly associated with.

Details

Each custom probeset identifier is mapped to a list of lists. The names on the outer list are GO identifiers. Each inner list contains of five named elements: GOID, Ontology, Term, Evidence and PMIDS.

The GOID element matches the GO identifier named in the outer list.

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 Term element indicates the name of GO cathegory corresponding to the GO identifier named in the outer list.

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

The PMIDS element indicates the Pubmed IDs of references in the literature, if relevant, obtained from EntrezGene.

NAs are assigned to probe identifiers that can not be mapped to any Gene Ontology information. Other information concerning Gene Ontology classification are available in a separate data package named GO.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert to a list
xx <- as.list(gahgu95eGO)
# 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"]]

}
```

---

gahgu95eINTERPRO   *Map between custom probeset Identifiers and Interpro domains*

Description

gahgu95eINTERPRO is an R environment that provides mappings between custom probeset identifiers and Interpro domains of corresponding protein.

Details

Each identifier is mapped to a named vector containing Interpro information about domains names. An NA is reported if there is no Interpro information for a given gene.

Each element name of named vector contains Interpro ID for corresponding domain.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eINTERPRO)
if(length(xx) > 0){
    # The Interpro domains for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eKEGG   *Map between custom probeset Identifiers and Kegg pathways*

Description

gahgu95eKEGG is an R environment that provides mappings between custom probeset identifiers and Kegg pathways information of corresponding genes.

Details

Each identifier is mapped to a named vector containing Kegg pathways names. An NA is reported if there is no Uniprot information for a given gene.

Within each vector, each element name indicate corresponding Kegg pathways ids.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eKEGG)
if(length(xx) > 0){
    # The Kegg pathways names for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eORGANISM   *The Organism for gahgu95e*

Description

gahgu95eORGANISM is an R environment that contains a single item: a character string that names the organism for which gahgu95e was built.

Details

Although the package name is suggestive of the organism for which it was built, gahgu95eORGANISM provides a simple way to programmatically extract the organism name.

Package built Mon Aug 25 17:16:58 2008

Examples

```
gahgu95eORGANISM
```

---

gahgu95e   *Bioconductor annotation data package*

Description

The annotation package was built using the following public data sources:

GeneCards database - www.genecards.org

---

gahgu95eSTRAND *Map custom probeset IDs to strand on chromosomes*

Description

gahgu95eSTRAND is an R environment that provides mappings between a custom probeset identifier and the strand of gene onto corresponding chromosome.

Details

Each identifier maps to a gene strand. If the gene strand is unknown, the vector will contain an NA.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built: Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert to a list
xx <- as.list(gahgu95eSTRAND)
# Remove probe identifiers that do not map to any strand
xx <- xx[!is.na(xx)]
if(length(xx) > 0){
    # Get the base pair number for gene strand for the first five probes
    xx[1:5]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eSWISSPROT *Map between custom probeset Identifiers and Swissprot IDs*

Description

gahgu95eSWISSPROT is an R environment that provides mappings between custom probeset identifiers and Swissprot IDs of corresponding proteins.

Details

Each identifier is mapped to a named vector containing Swissprot IDs. An NA is reported if there is no Uniprot information for a given gene.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eSWISSPROT)
if(length(xx) > 0){
    # The Swissprot IDs for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eSYMBOL   *Map between custom probeset Identifiers and Gene Symbols*

Description

gahgu95eSYMBOL is an R environment that provides mappings between custom probeset identifiers and gene abbreviations.

Details

Each 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:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

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

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eSYMBOL)
if(length(xx) > 0){
    # The symbols for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

---

gahgu95eUNIPROT   *Map between custom probeset Identifiers and Uniprot similarities and*
                         domains information

Description

gahgu95eUNIPROT is an R environment that provides mappings between custom probeset identi-
fiers and Interpro domains of corresponding protein.

Details

Each identifier is mapped to a named vector containing Uniprot information about similarities and
domains. An NA is reported if there is no Uniprot information for a given gene.

Each element name of named vector indicates wether corresponding element refers to similarity or
domain information from Uniprot.

Mappings were based on data provided by:

GeneCards database - www.genecards.org

Package built Mon Aug 25 17:16:58 2008

References

www.genecards.org

Examples

```
# Convert the environment to a list
xx <- as.list(gahgu95eUNIPROT)
if(length(xx) > 0){
    # The Uniprot domains or similarities for the first two elements of XX
    xx[1:2]
    # Get the first one
    xx[[1]]
}
```

Index