

# Package ‘HGNChelper’

March 21, 2014

**Maintainer** Levi Waldron <lwaldron.research@gmail.com>

**Depends** R (>= 2.10)

**Author** Levi Waldron and Markus Riester

**Version** 0.3.0

**Date** 2014-03-21

**License** GPL (>=2.0)

**Title** Handy functions for working with HGNC gene symbols and Affymetrix probeset identifiers.

**Description** Contains functions for  
identifying and correcting HGNC gene symbols which have been converted  
to date format by Excel, for reversibly converting between HGNC  
symbols and valid R names, identifying invalid HGNC symbols and  
correcting synonyms and outdated symbols which can be mapped to an official symbol.

**URL** <https://bitbucket.org/lwaldron/hgnchelper>

**BugReports** <https://bitbucket.org/lwaldron/hgnchelper/issues>

## R topics documented:

HGNChelper-package . . . . .	2
affyToR . . . . .	2
checkGeneSymbols . . . . .	3
findExcelGeneSymbols . . . . .	4
hgnc.table . . . . .	5
rToAffy . . . . .	5
rToSymbol . . . . .	6
symbolToR . . . . .	7
<b>Index</b>	<b>8</b>

---

HGNChelper-package	<i>Handy functions for working with HGNC gene symbols and Affymetrix probeset identifiers.</i>
--------------------	--

---

## Description

Contains functions for identifying and correcting HGNC gene symbols which have been converted to date format by Excel, for reversibly converting between HGNC symbols and valid R names, identifying invalid HGNC symbols and correcting synonyms and outdated symbols which can be mapped to an official symbol.

## Details

Package: HGNChelper  
 Maintainer: Levi Waldron <lwaldron@hsph.harvard.edu>  
 Depends: R (>= 2.10)  
 Author: Levi Waldron and Markus Riester  
 Version: 0.3.0  
 Date: 2014-03-21  
 License: GPL (>2.0)  
 Title: Handy functions for working with HGNC gene symbols and Affymetrix probeset identifiers.  
 URL: <https://bitbucket.org/lwaldron/hgnchelper>  
 BugReports: <https://bitbucket.org/lwaldron/hgnchelper>

## Author(s)

Levi Waldron and Markus Riester

---

affyToR	<i>function to convert Affymetrix probeset identifiers to valid R names</i>
---------	---

---

## Description

This function simply prepends "affy." to the probeset IDs to create valid R names. Reverse operation is done by the [rToAffy](#) function.

## Usage

```
affyToR(x)
```

## Arguments

x	vector of Affymetrix probeset identifiers, or any identifier which may with a digit.
---	--

**Value**

a character vector that is simply x with "affy." prepended to each value.

**Author(s)**

Levi Waldron and Markus Riester

---

checkGeneSymbols	<i>function to identify outdated or Excel-mogrified gene symbols</i>
------------------	--

---

**Description**

This function identifies gene symbols which are outdated or may have been mogrified by Excel or other spreadsheet programs. If output is assigned to a variable, it returns a data.frame of the same number of rows as the input, with a second column indicating whether the symbols are valid and a third column with a corrected gene list.

**Usage**

```
checkGeneSymbols(x, unmapped.as.na=TRUE, hgnc.table=NULL)
```

**Arguments**

x	Vector of gene symbols to check for mogrified or outdated values
unmapped.as.na	If TRUE, unmapped symbols will appear as NA in the Suggested.Symbol column. If FALSE, the original unmapped symbol will be kept when no correction can be found.
hgnc.table	If hgnc.table is a data.frame with colnames(hgnc.table) identical to c("Symbol", "Approved.Symbol"), it is used to correct gene symbols in x. Otherwise, the default table data("hgnc.table", package="HGNCHELPER") is used. The function used for creating this the hgnc.table dataframe can be found in the inst/hgncLookup.R file, in the source of this package. By default the hgnc.table dataframe shipped with this package is used (see ?hgnc.table)

**Value**

The function will return a data.frame of the same number of rows as the input, with corrections possible from hgnc.table.

**Author(s)**

Levi Waldron and Markus Riester

**See Also**

hgnc.table

## Examples

```
library(HGNChelper)

x = c("FN1", "TP53", "UNKNOWNGENE", "7-Sep", "9/7", "1-Mar", "Oct4", "4-Oct",
      "OCT4-PG4", "C19ORF71", "C19orf71")

res <- checkGeneSymbols(x)
res

if (interactive()){
  ##Run checkGeneSymbols with a brand-new map downloaded from HGNC:
  source(system.file("hgncLookup.R", package = "HGNChelper"))
  ##You should save this if you are going to use it multiple times,
  ##then load it from file rather than burdening HGNCs servers.
  ## save(hgnc.table, file="hgnc.table.rda", compress="bzip2")
  ## load("hgnc.table.rda")
  res <- checkGeneSymbols(x, hgnc.table=hgnc.table)
}
```

---

findExcelGeneSymbols    *function to identify Excel-mogrified gene symbols*

---

## Description

This function identifies gene symbols which may have been mogrified by Excel or other spreadsheet programs. If output is assigned to a variable, it returns a vector of the same length where symbols which could be mapped have been mapped. Note that this function is superceded by checkGeneSymbols, which corrects Excel-mogrified gene symbols as well as aliases and outdated symbols.

## Usage

```
findExcelGeneSymbols(x, mog.map = read.csv(system.file("extdata/mog_map.csv",
  package = "HGNChelper"), as.is = TRUE), regex =
"[0-9]\\-(JAN|FEB|MAR|APR|MAY|JUN|JUL|AUG|SEP|OCT|NOV|DEC)|[0-9]\\.[0-9]
[0-9]E\\+|[0-9][0-9]")
```

## Arguments

x	Vector of gene symbols to check for mogrified values
mog.map	Map of known mogrifications. A default map is available with this package by data(mog.map), but any map may be used. This should be a dataframe with two columns: original and mogrified, containing the correct and incorrect symbols, respectively.
regex	Regular expression, recognized by the base::grep function which is called with ignore.case=TRUE, to identify mogrified symbols. It is not necessary for all matches to have a corresponding entry in mog.map\$mogrified; values in x which are matched by this regex but are not found in mog.map\$mogrified simply will not be corrected. This regex is based that provided by Zeeberg et al., Mistaken Identifiers: Gene name errors can be introduced inadvertently when using Excel in bioinformatics. BMC Bioinformatics 2004, 5:80.

**Value**

if the return value of the function is assigned to a variable, the function will return a vector of the same length as the input, with corrections possible from `mog.map` made.

**Author(s)**

Levi Waldron and Markus Riester

**See Also**

`checkGeneSymbols`

---

hgnc.table	<i>All current and withdrawn HGNC symbols.</i>
------------	--

---

**Description**

All current and withdrawn symbols from [genenames.org](http://genenames.org).

**Format**

A dataframe with the first column providing a gene symbol or known alias (including withdrawn symbols), second column providing the approved HGNC symbol.

**Source**

Extracted from [genenames.org](http://genenames.org) prior to package build.

**Examples**

```
data("hgnc.table", package="HGNCHELPER", envir=environment())
head(hgnc.table)
```

---

rToAffy	<i>function to convert the output of affyToR back to the original Affymetrix probeset identifiers.</i>
---------	--

---

**Description**

This function simply strips the "affy." added by the [affyToR](#) function.

**Usage**

```
rToAffy(x)
```

**Arguments**

`x` the character vector returned by the `affyToR` function.

**Value**

a character vector of Affymetrix probeset identifiers.

**Author(s)**

Levi Waldron and Markus Riester

---

rToSymbol	<i>function to reverse the conversion made by symbolToR</i>
-----------	---

---

**Description**

This function reverses the actions of the symbolToR function.

**Usage**

```
rToSymbol(x)
```

**Arguments**

x                      the character vector returned by the symbolToR function.

**Value**

a character vector of HGNC gene symbols, which are not in general valid R names.

**Author(s)**

Levi Waldron and Markus Riester

**Examples**

```
library(HGNChelper)

data("hgnc.table", envir=environment())
hgnc.symbols <- as.character(na.omit(unique(hgnc.table[,2])))
if( !identical(all.equal(hgnc.symbols, rToSymbol(make.names(symbolToR(hgnc.symbols)))), TRUE))
  stop("HGNC mapping was not reversible.")
```

---

symbolToR	<i>function to *reversibly* convert HGNC gene symbols to valid R names.</i>
-----------	---

---

**Description**

This function reversibly converts HGNC gene symbols to valid R names by prepending "symbol.", and making the following substitutions: "-" to "hyphen", "@" to "ampersand", and "/" to "forward-slash".

**Usage**

```
symbolToR(x)
```

**Arguments**

x	vector of HGNC symbols
---	------------------------

**Value**

a vector of valid R names, of the same length as x, which can be converted to the same HGNC symbols using the rToSymbol function.

**Author(s)**

Levi Waldron and Markus Riester

**Examples**

```
library(HGNChelper)

data("hgnc.table", envir=environment())

hgnc.symbols <- as.character(na.omit(unique(hgnc.table[,2])))
if( !identical(all.equal(hgnc.symbols, rToSymbol(make.names(symbolToR(hgnc.symbols)))), TRUE))
  stop("HGNC mapping was not reversible.")
```

# Index

## \*Topic **datasets**

hgnc.table, [5](#)

## \*Topic **package**

HGNChelper-package, [2](#)

affyToR, [2](#), [5](#)

checkGeneSymbols, [3](#)

findExcelGeneSymbols, [4](#)

hgnc.table, [5](#)

HGNChelper-package, [2](#)

rToAffy, [2](#), [5](#)

rToSymbol, [6](#)

symbolToR, [7](#)