

Package ‘OncoScore’

October 12, 2016

Version 1.0.2

Date 2016-09-28

Title A tool to identify potentially oncogenic genes

Maintainer Daniele Ramazzotti <daniele.ramazzotti@yahoo.com>

Depends R (>= 3.3),

Imports biomaRt, grDevices, graphics, utils,

Suggests BiocGenerics, BiocStyle, testthat,

Description OncoScore is a tool to measure the association of genes to cancer based on citation frequency in biomedical literature. The score is evaluated from PubMed literature by dynamically updatable web queries.

Encoding UTF-8

LazyData TRUE

License GPL-3

URL <https://github.com/danro9685/OncoScore>

BugReports <https://github.com/danro9685/OncoScore>

biocViews BiomedicalInformatics

RoxygenNote 5.0.1

NeedsCompilation no

Author Daniele Ramazzotti [aut, cre],
Luca De Sano [aut],
Roberta Spinelli [ctb],
Carlo Gambacorti Passerini [ctb],
Rocco Piazza [ctb]

R topics documented:

compute.frequencies.scores	2
compute.oncoscore	3
compute.oncoscore.from.region	3
compute.oncoscore.timeseries	4

estimate.oncogenes	5
genes	5
get.genes.from.biomart	6
get.list.from.xml	7
get.pubmed.driver.analysis	7
perform.query	8
perform.query.timeseries	8
plot.oncoscore	9
plot.oncoscore.timeseries	10
query	11
query.timepoints	11
timepoints	12
try.scan	12
Index	13

compute.frequencies.scores
compute.frequencies.scores

Description

compute the logarithmic scores based on the frequencies of the genes

Usage

```
compute.frequencies.scores(data, filter.threshold = 1,
  analysis.mode = "Log2")
```

Arguments

`data` input data as result of the function `perform.query`
`filter.threshold` threshold to filter for a minimum number of citations for the genes
`analysis.mode` logarithmic scores to be computed, i.e., log10, log2, natural log or log5

Value

the computed scores

compute.oncoscore *compute.oncoscore*

Description

compute the OncoScore for a list of genes

Usage

```
compute.oncoscore(data, filter.threshold = 0, analysis.mode = "Log2",  
  cutoff.threshold = 21.09, file = NULL)
```

Arguments

data input data as result of the function perform.query
filter.threshold threshold to filter for a minimum number of citations for the genes
analysis.mode logarithmic scores to be computed, i.e., log10, log2, natural log or log5
cutoff.threshold threshold to be used to asses the oncogenes
file should I save the results to text files?

Value

the computed OncoScores and the clusters for the genes

Examples

```
data(query)  
compute.oncoscore(query)
```

compute.oncoscore.from.region
 compute.oncoscore.from.region

Description

Perform OncoScore analysis on a given chromosomal region

Usage

```
compute.oncoscore.from.region(chromosome, start = NA, end = NA,  
  gene.num.limit = 100, filter.threshold = NA, analysis.mode = "Log2",  
  cutoff.threshold = 21.09, file = NULL)
```

Arguments

chromosome	chromosome to be retrieved
start	initial position to be used
end	final position to be used
gene.num.limit	A limit to the genes to be considered in the analysis; this is done to limit the number of queries to PubMed
filter.threshold	threshold to filter for a minimum number of citations for the genes
analysis.mode	logarithmic scores to be computed, i.e., log10, log2, natural log or log5
cutoff.threshold	threshold to be used to asses the oncogenes
file	should I save the results to text files?

Value

the computed scores

Examples

```
chromosome = 15
start = 200000
end = 300000
compute.oncoscore.from.region(chromosome, start, end)
```

```
compute.oncoscore.timeseries
      compute.oncoscore.timeseries
```

Description

perform the OncoScore time series analysis for a list of genes and data times

Usage

```
compute.oncoscore.timeseries(data, filter.threshold = 0,
  analysis.mode = "Log2", cutoff.threshold = 21.09, file = NULL)
```

Arguments

data	input data as result of the function perform.query.timeseries
filter.threshold	threshold to filter for a minimum number of citations for the genes
analysis.mode	logarithmic scores to be computed, i.e., log10, log2, natural log or log5
cutoff.threshold	threshold to be used to asses the oncogenes
file	should I save the results to text files?

Value

the performed OncoScores time series analysis

Examples

```
data(query.timepoints)
compute.oncoscore.timeseries(query.timepoints)
```

estimate.oncogenes	<i>estimate.oncogenes</i>
--------------------	---------------------------

Description

estimate the oncoscore for the genes

Usage

```
estimate.oncogenes(data, cutoff.threshold = 21.09)
```

Arguments

data	input data as result of the function compute.frequencies.scores
cutoff.threshold	threshold to be used to asses the oncogenes

Value

the computed scores and oncogenes

genes	<i>A list of genes</i>
-------	------------------------

Description

This dataset contains a list of genes to be used in the analysis as an example

Usage

```
genes
```

Format

```
txt
```

Value

list of 5 elements

Source

example data

```
get.genes.from.biomart
```

```
get.genes.from.biomart
```

Description

Get a gene list from biomart

Usage

```
get.genes.from.biomart(chromosome, start = NA, end = NA)
```

Arguments

chromosome	chromosome to be retrieved
start	initial position to be used
end	final position to be used

Value

A list of genes

Examples

```
chromosome = 15  
start = 200000  
end = 300000  
ch15 = get.genes.from.biomart(chromosome, start, end)
```

`get.list.from.xml` *get.list.from.xml*

Description

process the result of the query

Usage

`get.list.from.xml(webget)`

Arguments

`webget` The result from the query to PubMed

Value

Processed result obtained from the query to PubMed

`get.pubmed.driver.analysis`
get.pubmed.driver.analysis

Description

query PubMed for a list of genes

Usage

`get.pubmed.driver.analysis(keywords, gene)`

Arguments

`keywords` The set of keywords to be used for the query to PubMed

`gene` The name of a gene to be used for the query to PubMed

Value

The frequency for the current gene retrieved with the query on the provided set of keywords

<code>perform.query</code>	<i>perform.query</i>
----------------------------	----------------------

Description

perform the query to PubMed

Usage

```
perform.query(list.of.genes, gene.num.limit = 100, custom.search = NA)
```

Arguments

`list.of.genes` The list of genes to be used in the queries to PubMed

`gene.num.limit` A limit to the genes to be considered in the analysis; this is done to limit the number of queries to PubMed

`custom.search` A custom set of keywords to be used when querying PubMed

Value

The frequencies of the genes in the cancer related documents and in all the documents retrieved on PubMed

Examples

```
data(genes)
perform.query(genes[1:2])
```

<code>perform.query.timeseries</code>	<i>perform.query.timeseries</i>
---------------------------------------	---------------------------------

Description

perform the query to PubMed for the time series analysis

Usage

```
perform.query.timeseries(list.of.genes, list.of.datatimes,
  gene.num.limit = 100, timepoints.limit = 10, custom.search = NA)
```


Arguments

list.of.genes The list of genes to be used in the queries to PubMed
list.of.datatimes The list of time points to be used in the queries to PubMed
gene.num.limit A limit to the genes to be considered in the analysis; this is done to limit the number of queries to PubMed
timepoints.limit A limit to the time points to be considered in the analysis; this is done to limit the number of queries to PubMed
custom.search A custom set of keywords to be used when querying PubMed

Value

The frequencies of the genes in the cancer related documents and in all the documents retrieved on PubMed at the specified time points

Examples

```
data(genes)
data(timepoints)
perform.query.timeseries(genes[1:2], timepoints[1:2])
```

plot.oncoscore *plot.oncoscore*

Description

plot the OncoScore for a list of genes

Usage

```
## S3 method for class 'oncoscore'
plot(x, gene.number = 5, main = "OncoScore",
     xlab = "score", ylab = "genes", file = NA, ...)
```

Arguments

x input data as result of the function compute.OncoScore
gene.number number of genes to print
main the title
xlab description of x axis (default score)
ylab description of y axis (default genes)
file where to save the plot
... additional parameter to pass to the barplot function

Value

A plot

Examples

```
data(query)
result = compute.oncoscore(query)
plot.oncoscore(result)
```

```
plot.oncoscore.timeseries
      plot.oncoscore.timeseries
```

Description

plot the OncoScore for a list of genes

Usage

```
## S3 method for class 'oncoscore.timeseries'
plot(x, gene.number = 5, incremental = FALSE,
     relative = FALSE, main = "OncoScore", xlab = "timepoints",
     ylab = "score", legend.pos = "top", file = NA, ...)
```

Arguments

<code>x</code>	input data as result of the function <code>compute.OncoScore</code>
<code>gene.number</code>	number of genes to print
<code>incremental</code>	display the OncoScore increment
<code>relative</code>	display the incrementa as relative value
<code>main</code>	the title
<code>xlab</code>	description of x asix (default score)
<code>ylab</code>	description of y asix (default genes)
<code>legend.pos</code>	Position of the legend
<code>file</code>	where to save the plot
<code>...</code>	additional parameter to pass to the lines function

Value

A plot

Examples

```
data(query.timepoints)
result = compute.oncoscore.timeseries(query.timepoints)
plot.oncoscore.timeseries(result)
```

query	<i>The result of perform.web.query on genes</i>
-------	-------------------------------------------------

Description

This dataset contains the result of perform.web.query on genes

Usage

```
query
```

Format

```
rdata
```

Value

```
matrix 5 x 2
```

Source

```
example data
```

query.timepoints	<i>The result of perform.time.series.query on genes and timepoints</i>
------------------	------------------------------------------------------------------------

Description

This dataset contains the result of perform.time.series.query on genes and timepoints

Usage

```
query.timepoints
```

Format

```
rdata
```

Value

```
list of 5 matrix 5 x 2
```

Source

example data

timepoints	<i>A list of timepoints</i>
------------	-----------------------------

Description

This dataset contains a list of time points to be used in the analysis as an example

Usage

timepoints

Format

txt

Value

list of 5 elements

Source

example data

try.scan	<i>try.scan</i>
----------	-----------------

Description

try to query the given URL

Usage

try.scan(getURL)

Arguments

getURL	The given URL
--------	---------------

Value

Result obtained from PubMed

Index

*Topic **datasets**

- genes, [5](#)
- query, [11](#)
- query.timepoints, [11](#)
- timepoints, [12](#)

- compute.frequencies.scores, [2](#)
- compute.oncoscore, [3](#)
- compute.oncoscore.from.region, [3](#)
- compute.oncoscore.timeseries, [4](#)

- estimate.oncogenes, [5](#)

- genes, [5](#)
- get.genes.from.biomart, [6](#)
- get.list.from.xml, [7](#)
- get.pubmed.driver.analysis, [7](#)

- perform.query, [8](#)
- perform.query.timeseries, [8](#)
- plot.oncoscore, [9](#)
- plot.oncoscore.timeseries, [10](#)

- query, [11](#)
- query.timepoints, [11](#)

- timepoints, [12](#)
- try.scan, [12](#)