

Package ‘jetset’

April 5, 2017

Version 3.4.0

Date 2017-04-05

Title One-to-One Gene-Probeset Mapping for Affymetrix Human
Microarrays

Author Qiyuan Li, Aron Eklund

Maintainer Aron Eklund <eklund@cbs.dtu.dk>

Depends R (>= 2.10)

Imports org.Hs.eg.db, AnnotationDbi

Description On Affymetrix gene expression microarrays, a single gene may be measured by multiple probe sets. This can present a mild conundrum when attempting to evaluate a gene “signature” that is defined by gene names rather than by specific probe sets. This package provides a one-to-one mapping from gene to “best” probe set for four Affymetrix human gene expression microarrays: hgu95av2, hgu133a, hgu133plus2, and u133x3p. This package also includes the pre-calculated probe set quality scores that were used to define the mapping.

License Artistic-2.0

LazyData Yes

URL <http://www.cbs.dtu.dk/biotools/jetset/>

NeedsCompilation no

Repository CRAN

Date/Publication 2017-04-05 11:15:11 UTC

R topics documented:

jmap	2
jscores	3
scores	4
Index	7

jmap

Retrieve jetset mapped probe sets

Description

This function retrieves probe sets corresponding to the queried genes

Usage

```
jmap(chip, eg, symbol, alias, ensembl)
```

Arguments

chip	Chip name
eg	A vector of Entrez GeneIDs (optional)
symbol	A vector of gene symbols (optional)
alias	A vector of gene aliases (optional)
ensembl	A vector of Ensembl IDs (optional)

Details

Currently, chip can be "hgu95av2", "hgu133a", "hgu133plus2", or "u133x3p".

Queried genes must be specified by either eg, symbol, alias, or ensembl.

If the query is not recognized, or is ambiguous, or corresponds to a gene that is not detected by the array, NA will be returned.

Details about the jetset algorithm are available in the vignette.

Value

A character vector of probe set IDs

References

Qiyuan Li, Nicolai J. Birkbak, Balazs Györfy, Zoltan Szallasi and Aron C. Eklund. (2011) Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*. 12:474.

See Also

The underlying Entrez ID to probeset data is available in (e.g.) `scores.hgu95av2`. Symbol, alias, and ensembl lookups are generated from e.g. `org.Hs.egSYMBOL2EG`.

Examples

```
genes <- c('MKI67', 'CHD5', 'ESR1', 'FGF19', 'ERBB2', 'NoSuchGene')

# This generates several informative warnings
jmap('hgu133a', symbol = genes)
```

jscores	<i>Retrieve jetset scores for probe sets</i>
---------	--

Description

This function retrieves jetset scores, which indicate the predicted quality of individual probe sets on selected Affymetrix microarrays.

Usage

```
jscores(chip, probeset, eg, symbol, alias, ensembl)
```

Arguments

chip	Chip name
probeset	A vector of probe set IDs (optional)
eg	A vector of Entrez GeneIDs (optional)
symbol	A vector of gene symbols (optional)
alias	A vector of gene aliases (optional)
ensembl	A vector of Ensembl IDs (optional)

Details

Currently, chip can be "hgu95av2", "hgu133a", "hgu133plus2", or "u133x3p". If no further arguments are specified, the scores for all probe sets on the chip are returned.

If any of probeset, eg, symbol, alias, or ensembl are specified, these are used to filter the resulting data frame in a logical OR sense.

Details about the jetset algorithm are available in the vignette.

Value

A data frame in which each row corresponds to a probe set, with 8 columns:

EntrezID	Entrez GeneID of the targeted gene (character).
nProbes	Number of probes in the probe set (integer).
process	Processivity requirement (integer).
specificity	Specificity score (numeric).
coverage	Coverage score (numeric).
robust	Robustness score (numeric).
overall	Overall score (numeric).
symbol	HUGO gene symbol (character).

The rows are sorted by decreasing overall score.

References

Qiyuan Li, Nicolai J. Birkbak, Balazs Gyorffy, Zoltan Szallasi and Aron C. Eklund. (2011) Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*. 12:474.

See Also

The underlying data comes from (e.g.) `scores.hgu95av2`, with gene symbol lookups coming from `org.Hs.egSYMBOL`.

Examples

```
genes <- c('MKI67', 'CHD5', 'ESR1', 'FGF19', 'ERBB2', 'NoSuchGene')

# This generates several informative warnings
jscores('hgu133a', symbol = genes)
```

scores

Data: Probe set quality scores

Description

This data set provides gene target and quality scores for each probe set on the corresponding Affymetrix gene expression microarrays.

Usage

```
scores.hgu95av2
scores.hgu133a
scores.hgu133plus2
scores.u133x3p
```

Format

A data frame with each row corresponding to a probe set, with 4 columns:

EntrezID Entrez GeneID of the targeted gene (character).

process Processivity requirement (integer).

specificity Specificity score (numeric).

coverage Coverage score (numeric).

Details

If there is a relative majority (plurality) of the probes in a probe set that are specific for a single gene, this is defined as the *targeted gene*. If no such majority exists, the targeted gene is defined as NA, as are the following scores.

The *processivity requirement* is the number of consecutive bases that must be synthesized to generate a target that can be detected by the probe set.

The *specificity score* is the fraction of the probes in a probe set that are likely to detect the targeted gene and unlikely to detect other genes.

The *coverage score* is the fraction of the splice isoforms belonging to the targeted gene that are detected by the probe set.

The following two scores are not contained in this data, but are calculated from the above scores; to see them use [jscores](#).

The *robustness score* quantifies robustness against transcript degradation. The robustness score uses the processivity requirement to estimate the signal intensity of a probe set, relative to the ideal case of perfect processivity.

The *overall score* is the product of the specificity score, coverage score, and robustness score.

All scores can range from 0 to 1. A higher score indicates better (predicted) performance.

Details about the jetset algorithm are available in the vignette.

Note

This data is also available in CSV format from <http://www.cbs.dtu.dk/biotools/jetset/>

Source

Scores are calculated from BLASTN alignments between probe sequences and Refseq transcript sequences, as described in the vignette and in the reference below.

The Refseq human RNA was downloaded from NCBI on 2017-04-04. The lookups were based on **org.Hs.eg.db** version 3.4.0.

References

Qiyuan Li, Nicolai J. Birkbak, Balazs Gyorffy, Zoltan Szallasi and Aron C. Eklund. (2011) Jetset: selecting the optimal microarray probe set to represent a gene. *BMC Bioinformatics*. 12:474.

See Also

[jscores](#) for a more convenient way to access this data

Examples

```
## Here is the EntrezID for the ESR1 gene
id <- "2099"

## Extract the scores for all probe sets detecting ESR1
scores.hgu95av2[which(scores.hgu95av2$EntrezID == id), ]
```

```
## Compare to the recommended function 'jscores'  
jscores("hgu95av2", eg = "2099")
```

Index

*Topic **datasets**

scores, [4](#)

*Topic **misc**

jmap, [2](#)

jscores, [3](#)

jmap, [2](#)

jscores, [3](#), [5](#)

org.Hs.egSYMBOL, [4](#)

org.Hs.egSYMBOL2EG, [2](#)

scores, [4](#)

scores.hgu95av2, [2](#), [4](#)