Package ‘cermt’

May 16, 2007

Version 0.0.9
Date 2007-04-12
Author Henning Redestig
Maintainer Henning Redestig <redestig@mpimp-golm.mpg.de>
Title CERMT Extracts Regulatory targets using Multiple Time courses
Depends R, methods, Biobase, pls, grid, limma
Description Main function can, given an AffyID of interest (e.g. a TF) and a set of multiple time series, extract a set of genes that are co-regulated (possibly time shifted) with the given TF in two or more treatments
License GPL version 2 (or later).
SaveImage no

R topics documented:

addGap .......................................................... 2
affy2agi ......................................................... 3
atgen ............................................................. 3
boxplotGap ...................................................... 4
cermt ............................................................ 4
cermtRes ......................................................... 5
gapRes .......................................................... 6
getAgi ............................................................. 7
getGap ............................................................. 7
getTargets ....................................................... 8
plotGap ........................................................... 8
plotRegulon ..................................................... 9
atgen ............................................................. 10

Index 11
addGap

Add Gap statistics to a cermtRes object

Description

Gap statistics can give an impression of how well the extracted regulon is separated from the other genes in terms of coherency and variance.

Usage

```r
addGap(crmRes, eset, r2star=NULL, numPerm=100, maxSize=1000, minSize=0, verbose=interactive(), ...)
```

Arguments

- `crmRes` a cermtRes object
- `eset` The expression set that was used to generate crmRes
- `r2star` Externally generated r2star
- `numPerm` The amount of resamples to be done. The more the better (and slower)
- `maxSize` The maximum genes to consider. Shorter means faster execution
- `minSize` The minimum genes to consider. CERMT can only detect fairly large regulons. Exactly how large they have to be depends on how strongly they respond and the total number of genes in the data set. Rule of thumb is that minimum regulon to search for must be larger 0.1% of the total amount.
- `verbose` Print a progress bar or not
- `...` Further arguments to the internal gap function such as r2 which can be either "cermt" or "hastie". "Hastie" is the clustering based Gap as described in Hastie et al. (2000)

Details

Gap statistics were introduced by Hastie et al. (2000) as a means of estimating appropriate cluster sizes. This method is an adaptation of the Gap to the CERMT problem. 'Regulon' are compared with how much better \( R^2 \) they have than what one would obtain when extracting with the regulator shuffled within the time series.

The CERMT R2 is simply the classical modelling R2 where we try to model the whole regulon from the expression of the transcription factor alone; \( R^2 = 1 - \sum \hat{X}^2 / \sum (X - \hat{X})^2 \).

Value

The cermtRes object with the gapData slot filled in.

Author(s)

Henning Redestig

See Also
**affy2agi**

Examples

**Description**

A simple mapping between Affymetrix ID’s and AGI codes.

Crude and dirty but kind of useful mapping. As obtained from TAIR.

**Usage**

data(affy2agi)

**Format**

A data frame

**Source**

http://www.arabidopsis.org

See Also

getAgi and getAffy

---

**atgen**

The abiotic stress series from AtGenExpress.

**Description**

Downloaded from TAIR, RMA normalized and averaged between the two technical replications. Genes with without a ’P’ call for at least two replica were excluded. Organellar encoded genes were also removed as well as genes with multiple hits for AGI codes.

**Usage**

data(atgen)

**Format**

An exprSet

**Source**

http://www.arabidopsis.org
boxplotGap

**Boxplot of Gap statistics**

**Description**
Visualize the Gap statistics using boxplots

**Usage**

```r
boxplotGap(x, r2star=NULL, ...)
```

**Arguments**

- `x`: A cermtRes object with Gap data
- `r2star`: A matrix with R2 statistics from a null distribution.
- `...`: Further arguments to `boxplot`

**Details**
Produces a plot with boxplots of the distribution of $R^2$ for each tested regulon size under the null hypothesis (the regulator can not extract a 'better' regulon than its shuffled version). The red points are the observed $R^2$ statistics for the unshuffled regulator (should be far away from the boxes for some regulon size.)

**Value**
None, used for side effect

**Author(s)**
Henning Redestig

**See Also**
- `plotGap`

---

cermt

**CERMT Extracts Regulatory targets using Multiple Time series**

**Description**
Given a putative regulator and an expression set with the regulator and all other genes measured in several short time series (with same amount of time points) then CERMT can extract (with the possibility of time shifts) a set of treatments in which the regulator is highly covariant with the same genes.
Usage

cermt(tf, eset, plsmethod = "onesidedoscores", verbose = TRUE,
  lags = 0:2, maxq2 = FALSE, method = c("cermt", "cor", "cov"),
  timepts = length(unique(eset$time)), checkMinDiff = TRUE,
  maxSize = 5000, alfa = 0.05, seedNum = 1,
  minFlat = NULL, minCont = NULL, atgen=TRUE)

Arguments

  tf        The index (name) of the regulator
  eset      The expression set to use. Must have components "treatment", "tissue" and
            "time" in its pData slot.
  plsmethod The PLS method to use
  verbose   Write some messages during processing or not
  lags      The lags to consider (0=no time shift, 1=genes responds to the regulator after
            one time point).
  maxq2     If a number then it is the amount of Q2 runs to do. If FALSE the correlation
            based method is used instead for determining the goodness-of-fit
  method    One of "cermt" "cor" or "cov". The latter two are just for comparison and are
            not recommended for use.
  timepts   The amount of measured time points
  checkMinDiff Exclude the treatments in which the regulator does not respond or not
  maxSize   The maxSize argument for addGap.
  alfa      The alfa value for the Q2 or correlation tests.
  seedNum   The rank of the best pair to use. 1 for the best pair, 2 for the second best pair etc.
  minFlat   The threshold for flat line
  minCont   The threshold for control similar
  atgen     Indicates if this is the atgenexpress dataset (as provided in this package) or not

Value

  A cermtRes object.

Author(s)

  Henning Redestig

---

cermtRes  

Class representation of CERMT statistics

Description

  This contains all the output from a CERMT analysis.
Creating Objects

\[
\text{new("cermtRes", stat=[statistic], used=[used treatments, lags], gapData=[gapRes object], method=[Used method], regulator=[Used regulator], geneNames=[Original gene names])}
\]

Slots

- **stat**: "numeric", The covariances between each gene and the regulator in the used treatments and lags.
- **used**: "list", A list with components 'cases'=the used treatments, 'caseNumber'=the original indices of the used treatments, 'lags'=the used lags.
- **gapData**: "gapRes", Gap statistic result
- **method**: "character",The used method.
- **regulator**: "character",The used regulator.
- **geneNames**: "character",The original gene names.

---

**gapRes**

*Class representation of Gap statistics*

Description

Usually part of a cermtRes object.

Creating Objects

\[
\text{new("gapRes", dstar=[the R2 values for the permuted regulator as a matrix], d=[the R2 values for the non-permuted regulator], use=[subsetting indices for each given size of the regulon], bestIndex=[The index of the best regulon size], bestSize=[The size of the best regulon], p=[The empirical p-value for the best regulon assesing the null-hypothesis that the regulon was obtained by a random regulator.])}
\]

Slots

- **dstar**: "matrix", The random R2 values.
- **d**: "numeric", The observed R2 values.
- **use**: "list", The indices defining the observed regulons
- **bestIndex**: "integer",The best index.
- **bestSize**: "integer",The size of the best regulon.
- **p**: "numeric",The empirical p-value.

See Also

- link{plotGap}, boxplotGap
getAgi

Convert between Affymetrix ID and AGI codes.

Description
Uses the affy2agi data to convert between the two different kinds of identifiers.

Usage
getAgi(affy, trim=TRUE)
getAffy(agi)

Arguments

affy    A set of affymetrix IDs
agi     A set of AGI codes
trim    Whether or not only the first nine characters of the AGI code should be returned

Details
Very slow method. Sorry about that.

Value

Author(s)
Henning Redestig

See Also
affy2agi

getGap

Get the Gap vector

Description
Get the Gap as a function of regulon size from a cermtRes object.

Usage
getGap(x, r2star)

Arguments

x   a cermtRes object
r2star   a matrix with pre-calculated r2star distribution
**getTargets**  
*Extract targets from a cermtRes object*

**Description**
Accessor function for extracting some targets from a cermtRes object.

**Usage**
```r
getTargets(cmRes, n = cmRes@gapData@bestSize)
```

**Arguments**
- `cmRes` A cermtRes object
- `n` The amount of genes to extract.

**Value**
A character vector with the extracted targets

**Author(s)**
Henning Redestig

---

**plotGap**  
*Visualize Gap statistics result.*

**Description**
Plot the Gap statistic as a function of regulon size.

**Usage**
```r
plotGap(x, r2star=NULL, log=TRUE,...)
```

**Arguments**
- `x` A cermtRes object
- `r2star` A matrix with R2 statistics from a null distribution.
- `log` Indicates if the x axis should be plotted on log scale or not
- `...` Further arguments to `plot`
**plotRegulon**

**Value**

None, used for its side effect.

**Author(s)**

Henning Redestig

**See Also**

boxplotGap

---

**Description**

Plot a group of genes as a regulon in each (or some) of the examined treatments. Based on the `coplot` function.

**Usage**

```r
plotRegulon(x, eset, tf = NULL, treatments = NULL, amount = 50,
            addLabels = TRUE, center = TRUE, scale = FALSE, ...)
```

**Arguments**

- `x`: Either a `cermtRes` object or character/numeric that defines as set of genes from `eset`.
- `eset`: The expression set used to generate `x`.
- `tf`: The index of the regulator. Not needed if `x` is a `cermtRes` object.
- `treatments`: The treatments to plot either as index to the vector obtained by `paste(eset$treatments, eset$tisse, sep=":")` or names therein. Not needed if `x` is a `cermtRes` object.
- `amount`: The amount of genes to plot.
- `addLabels`: Add labels to each plot or not.
- `center`: Center the genes within each treatment or not.
- `scale`: Scale the genes within each treatment or not to unit variance.
- `...`: Further arguments to `text` and `coplot`.

**Value**

None, used for its side effect.

**Author(s)**

Henning Redestig
atgen

The distribution of R2 under the null-hypothesis for abiotic stress series from AtGenExpress.

Description

More description to come.

Usage

data(r2star)

Format

A matrix

Source

Later
Index

*Topic **classes**
  cermtRes, 5
  gapRes, 6

*Topic **datasets**
  affy2agi, 3
  atgen, 3, 10

*Topic **multivariate**
  addGap, 1
  boxplotGap, 4
  cermt, 4
  getAgi, 7
  getGap, 7
  getTargets, 8
  plotGap, 8
  plotRegulon, 9

addGap, 1, 5
affy2agi, 3, 7
atgen, 3, 10

boxplotGap, 4, 6, 9

cermt, 4

cermtRes, 5, 5
cermtRes-class (cermtRes), 5

gapRes, 6
gapRes-class (gapRes), 6
getAffy, 3
getAffy (getAgi), 7
getAgi, 3, 7
getGap, 7
getTargets, 8

plotGap, 4, 8
plotRegulon, 9

r2star (atgen), 10

show, cermtRes-method (cermtRes), 5