Package 'GSAQ'

July 21, 2025

Type Package

Title Gene Set Analysis with QTL

Version 1.0

Date 2016-08-26

Author Samarendra Das <samarendra.das@icar.gov.in>

Maintainer Samarendra Das <samarendra.das@icar.gov.in>

Depends R (>= 3.3.1)

Description Computation of Quantitative Trait Loci hits in the selected gene set. Perform-
ing gene set validation with Quantitative Trait Loci information. Performing gene set enrich-
ment analysis with available Quantitative Trait Loci data and computation of statistical signifi-
cance value from gene set analysis. Obtaining the list of Quantita-
tive Trait Loci hit genes along with their overlapped Quantitative Trait Loci names.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2016-09-08 13:07:11

Contents

---

genedist   *Chromosomal distribution of the genes in the selected geneset*

Description

The function computes the chromosome wise distribution of the genes in the selected geneset and also plots the chromosomal distribution.

Usage

```
genedist(geneset, genelist, plot)
```

Arguments

geneset   geneset is a vector of characters representing the names of genes/ gene ids selected from the whole gene list by using a gene selection method.

genelist   genelist is a N by 3 dataframe/ matrix (genes/gene ids as row names); where, N represents the number of genes in the whole gene space: first coloumn represnting the chromosomal location of genes: second coloumn representing the start position of genes in terms of basepairs: third coloumn representing the end position of genes in terms of basepairs in their respective chromosomes.

plot   plot is a character string indicating whether the chromosomal distribution of the genes in the selected geneset will be plotted or not. It can be either TRUE/FALSE.

Value

The function returns the chromosomal distribution of the genes in the selected geneset.

Author(s)

Samarendra Das

Examples

```
data(rice_salt)
data(genelist)
x=as.data.frame(rice_salt[-1,])
y=as.numeric(rice_salt[1,])
geneset= GeneSelect(x, y, s=50, method="t-score")$selectgenes
genelist=as.data.frame(genelist)
genedist(geneset, genelist, plot=TRUE)
```

---

genelist	*A list of genes of rice*

Description

This data is in form of a 200 by 3 dataframe with genes/gene ids as rownames. The first column represents the chromosomal location of the genes (chromosome number). The second coloumn represents start position of the genes in terms of basepairs (bps) and the third coloumn represents end position of genes in terms of basepairs (bps) in their respective chromosomes.

Usage

```
data("genelist")
```

Format

A data frame with 200 rows as genes and the columns represent the chromosomal locations, start positions and end positions of respective genes.

Chr chr represents the chromosomal location of the genes

Start start represents the start position of the genes in their respective chromosomes

End End represents the end position of the genes in their respective chromosomes

Details

The data is created by taking 200 genes from the large number of genes from NCBI GEO database. The genomic location of the genes on the rice genome are obtained from MSU Rice Genome Annotation (Osa1).

Source

Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.ncbi.nlm.nih.gov/geo/. Ouyang S, Zhu W, Hamilton J, Lin H, Campbell M, et al. (2007) The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Research 35.

Examples

```
data(genelist)
```

---

geneqtl	*List of the selected genes along with their corresponding overlapped QTL*

Description

The function enables to obtain list of the selected genes along with the corresponding overlapped Quantitative Trait Loci (QTL) ids/names along with their genomic positions.

Usage

```
geneqtl(geneset, genelist, qtl)
```

Arguments

geneset   geneset is a vector of characters representing the names of genes/ gene ids se-
                lected from the whole gene list/space by using a gene selection method.

genelist   genelist is a N by 3 dataframe/ matrix (genes/gene ids as row names); where, N
                represents the number of genes in the whole gene set: first coloumn represnting
                the chromosomal location of genes: second coloumn representing the start posi-
                tion of genes in terms of basepairs: third coloumn representing the end position
                of genes in terms of basepairs in their respective chromosomes.

qtl   qtl is a Q by 3 dataframe/matrix (qtl names/qtl ids as row names);where, Q rep-
                resents the number of qtls: first coloumn represnting the chromosomal location
                of qtls: second coloumn representing the start position of qtls in terms of base-
                pairs: third coloumn representing the end position of qtls in terms of basepairs
                in their respective chromosomes.

Value

The function returns a list with two components. First component returns the list of selected genes
along with their overlapped QTL ids/names. Second component gives the list of selected genes with
their overlapped QTL ids/names and their respective genomic positions.

Author(s)

Samarendra Das

Examples

```
data(rice_salt)
data(genelist)
data(qtl_salt)
x=as.data.frame(rice_salt[-1,])
y=as.numeric(rice_salt[1,])
genelist=as.data.frame(genelist)
qtl=as.data.frame(qtl_salt)
geneset= GeneSelect(x, y, s=50, method="t-score")$selectgenes
```

```
geneqtl(geneset, genelist, qtl)
```

GeneSelect   *Selection of informative geneset*

Description

The function returns the informative geneset from the high dimensional gene expression data using a proper statistical technique.

Usage

```
GeneSelect(x, y, s, method)
```

Arguments

x	x is a N x m gene expression data matrix (must be data frame) and row names as gene names, where, N represents the number of genes in the whole gene space and m is number of samples.
y	y is a m by 1 vector representing the sample labels, is according to the different stress conditions for two class problem (must be 1: stress/-1: control)
s	s is a numeric constant representing the number of genes to be selected from the large pool of genes/ gene space.
method	method is a character string indicating which method for informative gene selection is to be used. One of method "t-score" (default), "F-score", "MRMR", "BootMRMR" can be abbreviated and used.

Value

The function returns the informative geneset using a particular method from the high dimensional gene expression data.

Author(s)

Samarendra Das

Examples

```
data(rice_salt)
x=as.data.frame(rice_salt[-1,])
y=as.numeric(rice_salt[1,])
GeneSelect(x, y, s=50, method="t-score")$selectgenes
```

GSAQ	*Gene Set Analysis with Quantitative Trait Loci with gene sampling model*

Description

The function computes the statistical significance value (p-value) from gene set analysis test with QTL for the test H0: Genes in the selected geneset are at most as often overlapped with the QTL regions as the genes in not selected geneset; against H1: Genes in the geneset are more often overlapped with the QTL regions as compared to genes in not selected geneset.

Usage

```
GSAQ(geneset, genelist, qtl, SampleSize, K, method)
```

Arguments

geneset   geneset is a vector of characters representing the names of genes/ gene ids selected from the whole gene list by using a gene selection method.

genelist   genelist is a N by 3 dataframe/ matrix (genes/gene ids as row names); where, N represents the number of genes in the whole gene space: first coloumn represnting the chromosomal location of genes: second coloumn representing the start position of genes in terms of basepairs: third coloumn representing the end position of genes in terms of basepairs in their respective chromosomes.

qtl   qtl is a Q by 3 dataframe/matrix (qtl names/qtl ids as row names);where, Q represents the number of qtls: first coloumn represnting the chromosomal location of qtls: second coloumn representing the start position of qtls in terms of basepairs: third coloumn representing the end position of qtls in terms of basepairs in their respective chromosomes.

SampleSize   SampleSize is a numeric constant representing the size of the gene sample drawn from the geneset using the gene sampling model (SampleSize must be less than the size of geneset).

K   K is a numeric constant representing the number of gene samples of size equal to SampleSize will be drawn by the using gene sampling model.

method   method is a character string indicating which method for final p-value (combining p-values for various gene samples) is to be computed. One of "meanp", "sump", "logit", "sumz"or "logp" (default) can be abbreviated and used.

Value

The function returns the final statistical significance value (p-value) from Gene set Analysis with QTL test.

Author(s)

Samarendra Das

Examples

```
data(rice_salt)
data(genelist)
data(qtl_salt)
x=as.data.frame(rice_salt[-1,])
y=as.numeric(rice_salt[1,])
genelist=as.data.frame(genelist)
qtl=as.data.frame(qtl_salt)
geneset= GeneSelect(x, y, s=50, method="t-score")$selectgenes
GSAQ(geneset, genelist, qtl, SampleSize=30, K=50, method="meanp")
```

GSVQ	*Gene Set Validation with QTL using Hyper-geometric test without gene sampling model*

Description

The function computes ths statisical significance value (p-value) for gene set validation using hypergeometric test.

Usage

```
GSVQ(geneset, genelist, qtl)
```

Arguments

geneset
: geneset is a vector of characters representing the names of genes/ gene ids selected from the whole gene list by using a gene selection method.

genelist
: genelist is a N by 3 dataframe/ matrix (genes/gene ids as row names): where, N represents the number of genes in the whole gene set: first coloumn represnting the chromosomal location of genes: second coloumn representing the start position of genes in terms of basepairs: third coloumn representing the end position of genes in terms of basepairs.

qtl
: qtl is a Q by 3 dataframe/matrix (qtl names/qtl ids as row names);where, Q represents the number of qtls: first coloumn represnting the chromosomal location of qtls: second coloumn representing the start position of qtls in terms of basepairs: third coloumn representing the end position of qtls in terms of basepairs.

Value

The function returns the statisical significance value (p-value) from Hyper-geometric test for validation of the selected gene set with qtl data.

Author(s)

Samarendra Das

Examples

```
data(rice_salt)
data(genelist)
data(qtl_salt)
x=as.data.frame(rice_salt[-1,])
y=as.numeric(rice_salt[1,])
geneset= GeneSelect(x, y, s=50, method="t-score")$selectgenes
genelist=as.data.frame(genelist)
qtl=as.data.frame(qtl_salt)
GSVQ(geneset, genelist, qtl)
```

---

qtldist   *QTL wise distribution of genes in the selected geneset*

Description

Computation of number of qtl-hit genes in each QTL and also QTL wise distribution of genes in the selected geneset

Usage

```
qtldist(geneset, genelist, qtl, plot)
```

Arguments

geneset	geneset is a vector of characters representing the names of genes/ gene ids selected from the whole gene space by using a gene selection method.
genelist	genelist is a N by 3 dataframe/ matrix (genes/gene ids as row names); where, N represents the number of genes in the whole gene space: first coloumn represnting the chromosomal location of genes: second coloumn representing the start position of genes in terms of basepairs: third coloumn representing the end position of genes in terms of basepairs in their respective chromosomes.
qtl	qtl is a Q by 3 dataframe/matrix (qtl names/qtl ids as row names);where, Q represents the number of qtls: first coloumn represnting the chromosomal location of qtls: second coloumn representing the start position of qtls in terms of basepairs: third coloumn representing the end position of qtls in terms of basepairs in their respective chromosomes.
plot	plot is a character string used to plot the QTL wise distribution of genes in the selected gene set. It can be either TRUE/FALSE.

Value

The function returns number of qtl-hit genes in each QTL and QTL wise distribution of the selected genes.

Author(s)

Samarendra Das

Examples

```
data(rice_salt)
data(genelist)
data(qtl_salt)
x=as.data.frame(rice_salt[-1,])
y=as.numeric(rice_salt[1,])
geneset= GeneSelect(x, y, s=50, method="t-score")$selectgenes
genelist=as.data.frame(genelist)
qtl=as.data.frame(qtl_salt)
qtldist(geneset, genelist, qtl, plot=TRUE)
```

---

qtlhit   *Computation of qtl-hit statistic for the selected gene set*

Description

The function computes the statistic, i.e. number of qtl-hit genes in the selected gene set.

Usage

```
qtlhit(geneset, genelist, qtl)
```

Arguments

geneset	geneset is a vector of characters representing the names of genes/ gene ids selected from the whole gene list by using a gene selection method.
genelist	genelist is a N by 3 dataframe/ matrix (genes/gene ids as row names); where, N represents the number of genes in the whole gene set: first coloumn represnting the chromosomal location of genes: second coloumn representing the start position of genes in terms of basepairs: third coloumn representing the end position of genes in terms of basepairs.
qtl	qtl is a Q by 3 dataframe/matrix (qtl names/qtl ids as row names);where, Q represents the number of qtls: first coloumn represnting the chromosomal location of qtls: second coloumn representing the start position of qtls in terms of basepairs: third coloumn representing the end position of qtls in terms of basepairs.

Value

The function returns a numeric value of the statistic 'qtl-hit' representing the number of qtl-hits by the genes in the selected gene set.

Author(s)

Samarendra Das

Examples

```
data(rice_salt)
data(genelist)
data(qtl_salt)
x=as.data.frame(rice_salt[-1,])
y=as.numeric(rice_salt[1,])
geneset= GeneSelect(x, y, s=50, method="t-score")$selectgenes
genelist=as.data.frame(genelist)
qtl=as.data.frame(qtl_salt)
qtlhit(geneset, genelist, qtl)
```

---

qtl_salt   *A list of salt responsive Quantitative Trait Loci of rice*

Description

This data is in form of a 13 by 3 dataframe with qtls/qtl ids as rownames. The first column reore-
sents the chromosomal location of the respective qtls (chromosome number). The second coloumn
represents start position of the qtls in terms of basepairs (bps) and the third coloumn represents end
position of qtls in terms of basepairs (bps) in their respective chromosomes.

Usage

```
data("qtl_salt")
```

Format

A data frame with 13 rows as qtl and the columns represent the chromosomal locations, start posi-
tions and end positions of respective qtls.

Chr chr represents the chromosomal location of the qtls

Start start represents the start position of the qtls in their respective chromosomes

End End represents the end position of the qtls in their respective chromosomes

Details

The data is created by taking 13 unique salt responsive qtls from the Gramene QTL database. The
genomic locations of these QTLs on rice genome are obtained using Gramene annotation of MSU
Rice Genome Annotation (Osa1).

Source

Gramene QTL library (http://www.gramene.org/qtl/). Ouyang S, Zhu W, Hamilton J, Lin H, Camp-
bell M, et al. (2007) The TIGR Rice Genome Annotation Resource: improvements and new fea-
tures. Nucleic Acids Research 35: D883-D887.

Examples

```
data(qtl_salt)
```

---

rice_salt   *Gene expression data of rice under salinity stress*

Description

This data has gene expression values of 200 genes over 40 microarray samples/subjects for a salinity vs. control study in rice. These 40 samples belong to either of salinity stress or control condition (two class problem). This gene expression data is balanced type as the first 20 samples are under salinity stress and the later 20 are under control condition. The first row of the data contains the samples/subjects labels with entries are 1 and -1, where the labels '1' and '-1' represent samples generated under salinity stress and control condition respectively.

Usage

```
data("rice_salt")
```

Format

A data frame with 200 genes over 40 microarray samples/subjects.

Details

The data is created by taking 200 genes from the large number of genes from NCBI GEO database. The rows are the genes and columns are the samples/subjects. The first half of the samples/subjects are generated under salinity stress condition and other half under control condition.The first row of the data contains the samples/subjects labels with entries as 1 and -1, where th label '1' and '-1' represents sample generated under salinity stress and control condition respectively.

Source

Gene Expression Omnibus: NCBI gene expression and hybridization array data repository.ncbi.nlm.nih.gov/geo/.

Examples

```
data(rice_salt)
```

---

totqtlhit   *Computation of total number of qtl-hits found in the whole gene space*

Description

It enable to Compute the total number qtl-hits found in the whole gene space or in the micro-array chip

Usage

```
totqtlhit(genelist, qtl)
```

Arguments

genelist   genelist is a N by 3 dataframe/ matrix (genes/gene ids as row names); where, N
                  represents the number of genes in the whole gene set: first coloumn represnting
                  the chromosomal location of genes: second coloumn representing the start posi-
                  tion of genes in terms of basepairs: third coloumn representing the end position
                  of genes in terms of basepairs.

qtl   qtl is a Q by 3 dataframe/matrix (qtl names/qtl ids as row names);where, Q rep-
                  resents the number of qtls: first coloumn represnting the chromosomal location
                  of qtls: second coloumn representing the start position of qtls in terms of base-
                  pairs: third coloumn representing the end position of qtls in terms of basepairs.

Value

The function returns a numeric value representing the total number of qtl-hits found in the whole
gene list or in a micro-array chip.

Author(s)

Samarendra Das

Examples

```
data(genelist)
data(qtl_salt)
genelist=as.data.frame(genelist)
qtl=as.data.frame(qtl_salt)
totqtlhit(genelist, qtl)
```

Index