Package ‘traseR’

April 23, 2016

Type Package
Title GWAS trait-associated SNP enrichment analyses in genomic intervals
Version 1.0.0
Depends R (>= 3.2.0), GenomicRanges, IRanges, BSgenome.Hsapiens.UCSC.hg19
Suggests BiocStyle, RUnit, BiocGenerics
Date 2015-10-13
Author Li Chen, Zhaohui S. Qin
Maintainer li.chen@emory.edu
Description traseR performs GWAS trait-associated SNP enrichment analyses in genomic intervals using different hypothesis testing approaches, also provides various functionalities to explore and visualize the results.
License GPL
LazyLoad yes
biocViews Genetics, Sequencing, Coverage, Alignment, QualityControl, DataImport
NeedsCompilation no

R topics documented:

traseR-package ......................................................... 2
CEU ................................................................. 2
plots ................................................................. 3
print.traseR ........................................................ 4
queries .............................................................. 5
taSNPDB ................................................................ 6
taSNPLDDB .......................................................... 7
Tcell ..................................................................... 8
traseR ................................................................. 8

Index 11
**traseR-package**

*GWAS trait-associated SNP enrichment analyses in genomic intervals*

**Description**

Perform GWAS trait-associated SNP enrichment analyses in genomic intervals. Explore and visualize the results.

**Details**

<table>
<thead>
<tr>
<th>Package</th>
<th>traseR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Package</td>
</tr>
<tr>
<td>Version</td>
<td>1.0</td>
</tr>
<tr>
<td>Date</td>
<td>2015-10-08</td>
</tr>
<tr>
<td>License</td>
<td>GPL</td>
</tr>
</tbody>
</table>

**Author(s)**

Li Chen <li.chen@emory.edu>, Zhaohui S.Qin<zhaohui.qin@emory.edu>

---

**ceu**

*Sampled SNPs from all SNPs of CEU population*

**Description**

A GRRange object `ceu` contains 5% of all SNPs from CEU by controlling genome-wide density is the same as all SNPs from CEU.

**Usage**

data(ceu)

**Value**

The data frame `ceu` contains three columns,

- **SNP_ID**: SNP rs number
- **seqnames**: Chromosome number associated with rs number
- **ranges**: Chromosomal position, in base pairs, associated with rs number

**Author(s)**

Li Chen <li.chen@emory.edu>, Zhaohui S.Qin<zhaohui.qin@emory.edu>
Visualize of trait-associated SNPs

Description

These are a group of functions to generate plots to visualize the trait-associated SNPs.

Usage

plotContext(snpdbL region=NULLL keyword = NULLL pvalue = 1e-3)
plotPvalue(snpdb, region=NULLL keyword = NULLL plot.type = c("densityplot", "boxplot"), pvalue = 1e-3)
plotSNP(snpdb, snpid, ext = 10000)
plotGene(snpdb, gene, ext = 10000)
plotInterval(snpdb, interval, ext = 10000)

Arguments

snpdb A GRange object or data frame, which is GWAS trait-associated SNPs downloaded from up-to-date dbGaP and NHGRI public database. It is maintained to be updated to the latest version. The data frame contains the following columns, Trait, SNP, p-value, Chr, Position, Context, GENE_NAME, GENE_START, GENE_END, GENE_STRAND. The data frame is in data subdirectory. Users are free to add more SNP records to the data frame for practical use.

region A data frame, which is genomic intervals with three columns, chromosome, genomic start position, genomic end position.

keyword The keyword is used when specific trait is of interest. If keyword is specified, only the SNPs associated to the trait are used for analyses. Otherwise, all traits will be analyzed.

snpid SNP rs number

gene Gene name

pvalue SNPs with p-value less than this threshold are used for analyses.

plot.type Either "densityplot" or "boxplot"

ext Bp extended upstream and downstream

xymax The maximum range on x-axis and y-axis

interval A data frame, genomic interval: chromosome, genomic start position, genomic end position
Value

<table>
<thead>
<tr>
<th>Function</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>plotContext</td>
<td>A pie plot with the distribution of SNP function class</td>
</tr>
<tr>
<td>plotPvalue</td>
<td>A density plot of -logPvalue of trait-associated SNPs</td>
</tr>
<tr>
<td>plotSNP</td>
<td>A plot of trait-associated SNP on chromosome</td>
</tr>
<tr>
<td>plotGene</td>
<td>A plot with the gene and possible nearby trait-associated SNPs</td>
</tr>
<tr>
<td>plotInterval</td>
<td>A plot with chromosome interval with possible nearby genes and trait-associated SNPs</td>
</tr>
</tbody>
</table>

Author(s)

Li Chen <li.chen@emory.edu>, Zhaohui Qin <zhaohui.qin@emory.edu>

Examples

```r
data(tasnpdb)
plotContext(snpdb = tasnpdb, keyword = "Autoimmune")
plotGene(snpdb = tasnpdb, gene = "ZFP92", ext = 50000)
plotSNP(snpdb = tasnpdb, snpid = "rs766420", ext = 50000)
plotInterval(snpdb = tasnpdb, data.frame(chr = "chrX", start = 152633780, end = 152737085))
```

Description

Print the outcome of taSNP enrichment analyses. Print the overall taSNP enrichment and trait-specific taSNP enrichment.

Usage

```r
## S3 method for class 'traseR'
print(x, isTopK = FALSE, topK = 10, trait.threshold = 10, ...)
```

Arguments

- `x` Object returned from `traseR`
- `isTopK` If `isTopK` is TRUE, topK traits are printed; otherwise, traits with p-value below Bonferroni correction threshold are printed. Default is FALSE.
- `topK` Top K traits are printed. Default is 10.
- `trait.threshold` Traits above this threshold are reported. Default is 10.
- `...` Other parameters to print
Value

Print a data frame of traits ranked by p-value

Author(s)

Li Chen <li.chen@emory.edu>, Zhaohui S. Qin <zhaohui.qin@emory.edu>

Examples

```r
data(taSNPDB)
data(Tcell)
x = traser(snpdb = taSNPDB, region = Tcell)
print(x)
```

Here is a small snippet of the content:

---

**Value**

Print a data frame of traits ranked by p-value

**Author(s)**

Li Chen <li.chen@emory.edu>, Zhaohui S. Qin <zhaohui.qin@emory.edu>

**Examples**

```r
data(taSNPDB)
data(Tcell)
x = traser(snpdb = taSNPDB, region = Tcell)
print(x)
```

**Description**

These are a group of functions to retrieve the trait-associated SNPs based on input

**Usage**

```r
queryKeyword(snpdb, region = NULL, keyword = NULL, returnby = c("SNP_ID", "trait"), pvalue = 1e-3)
queryGene(snpdb, genes = NULL)
querySNP(snpdb, snpid, region = NULL)
```

**Arguments**

- `snpdb`: A GRrange object or data frame, which is GWAS trait-associated SNPs downloaded from up-to-date dbGaP and NHGRI public database. It is maintained to be updated to the latest version. The data frame contains the following columns, Trait, SNP_ID, p.value, Chr, Position, Context, GENE_NAME, GENE_START, GENE_END, GENE_STRAND. The data frame is in data subdirectory. Users are free to add more SNP records to the data frame for practical use.

- `region`: A data frame, which is genomic intervals with three columns, chromosome, genomic start position, genomic end position.

- `keyword`: The keyword is used when specific trait is of interest. If keyword is specified, only the SNPs associated to the trait are used for analyses. Otherwise, all traits will be analyzed.

- `snpid`: SNP rs number

- `genes`: Gene name
pvalue SNPs with p-value less than this threshold are used for analyses.
returnby Either SNP or trait. If returnby is specified as 'SNP_ID', a data frame based on 'SNP_ID' is returned. If returnby is specified as 'trait', a data frame based on 'trait' is returned.

Value
queryKeyword: Return a data frame of traits containing the keyword queryGene: Return a data frame of traits associated with the gene querySNP: Return a data frame of traits associated with the SNP

Author(s)
Li Chen <li.chen@emory.edu>, Zhaohui Qin<zhaohui.qin@emory.edu>

Examples

data(tasnpdb)
data(Tcell)
x=queryKeyword(snpdb=tasnpdb,region=Tcell,keyword="Autoimmune",returnby="SNP_ID")
x=queryGene(snpdb=tasnpdb,genes=c("AGRN","UBE2J2","SSU72"))
x=querySNP(snpdb=tasnpdb,snpid=c("rs3766178","rs880051"))

---

Description
A GRang object tasnpdb contains trait-associated SNPs from dbGaP and NHGRI downloaded from Association Results Browser.

Usage
data(tasnpdb)

Value
The data frame tasnpdb contains the following columns

<table>
<thead>
<tr>
<th>Trait</th>
<th>Trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP_ID</td>
<td>SNP rs number</td>
</tr>
<tr>
<td>p.value</td>
<td>GWAS SNP p-value</td>
</tr>
<tr>
<td>seqnames</td>
<td>Chromosome</td>
</tr>
<tr>
<td>ranges</td>
<td>Chromosome position</td>
</tr>
</tbody>
</table>
tasnplldb

<table>
<thead>
<tr>
<th>Context</th>
<th>SNP functional class</th>
</tr>
</thead>
<tbody>
<tr>
<td>GENE_NAME</td>
<td>Nearest gene name</td>
</tr>
<tr>
<td>GENE_START</td>
<td>Gene start genomic position</td>
</tr>
<tr>
<td>GENE_END</td>
<td>Gene end genomic position</td>
</tr>
<tr>
<td>GENE_STRAND</td>
<td>Gene strand</td>
</tr>
</tbody>
</table>

**Author(s)**

Li Chen <li.chen@emory.edu>, Zhaohui S.Qin<zhaohui.qin@emory.edu>

---

**Description**

A GRRange object tasnplldb contains linkage disequilibrium (>0.8) SNPs of all trait-associated SNPs from dbGaP and NHGRI.

**Usage**

data(tasnplldb)

**Value**

The data frame tasnplldb contains four columns,

- **SNP_ID**: SNP rs number
- **seqnames**: Chromosome number associated with rs number
- **ranges**: Chromosomal position, in base pairs, associated with rs number
- **Trait**: Trait the SNP is associated with

**Author(s)**

Li Chen <li.chen@emory.edu>, Zhaohui S.Qin<zhaohui.qin@emory.edu>
Tcell  
*Peak regions of H3K4me1 in Peripheral blood T cell*

**Description**
A GRange object Tcell contains three columns: chromosome, genomic start position and genomic end position.

**Usage**
data(Tcell)

**Value**
The data frame Tcell contains three columns,

- **seqnames**: Chromosome id
- **ranges**: Chromosome position

**Author(s)**
Li Chen <li.chen@emory.edu>, Zhaohui S.Qin <zhaohui.qin@emory.edu>

---

traseR  
*TRait-Associated SNP EnRichment analyses*

**Description**
Perform GWAS trait-associated SNP enrichment analyses in genomic intervals using different approaches.

**Usage**
traseR(snpdb, region, snpdb.bg=NULL, keyword = NULL, rankby = c("pvalue","odds.ratio"), test.method = c("binomial","fisher","chisq","nonparametric"), alternative = c("greater","less","two.sided"), ntimes=100,nbatch=1, trait.threshold = 0, pvalue = 1e-3)
Arguments

snpdb  A GRange object. It could be GWAS trait-associated SNPs downloaded from up-to-date dbGaP and NHGRI public database. It is maintained to be updated to the latest version. The data frame contains the following columns, Source, Trait, SNP, p.value, Chr, Position. The data frame is in data subdirectory. Users are free to add more SNP records to the data frame for practical use. It could also be a data frame with columns as, SNP, Chr, Position.

region  A GRange object or data frame, which is genomic intervals with three columns, chromosome, genomic start position, genomic end position.

snpdb.bg  A GRange object contains non-trait-associated SNPs. They are treated as background for statistical testing instead of whole genome as background if specified.

keyword  The keyword is used when specific trait is of interest. If keyword is specified, only the SNPs associated to the trait are used for analyses. Otherwise, all traits will be analyzed.

rankby  Traits could be ranked by either p-value or adds.ratio based on the enrichment level of trait-associated SNPs in genomic intervals.

test.method  Several hypothesis testing options are provided: binomial(binomial test), fisher(Fisher's exact test), chisq(Chi-squared test), chisq(nonparametric test). Default is binomial(binomial test)

alternative  Indicate the alternative hypothesis. If greater, test if the genomic intervals are enriched in trait-associated SNPs than background. If less, test if the genomic intervals are depleted in trait-associated SNPs than background. If two.sided, test if there is difference between the enrichment of trait-associated SNPs in genomic intervals and in background.

ntimes  The number of shuffling time for one batch. See nbatch.

nbatch  The number of batches. The product of ntimes and nbatch is the total number of shuffling time.

trait.threshold  Test traits with number of SNPs more than the threshold.

pvalue  SNPs with p-value less than this threshold are used for analyses.

Details

Return a list that contains two data frames and one integer. One data frame tb.all contains the results of enrichment analyses for all trait-associated SNPs in genomic intervals. Another data frame tb contains the results of enrichment analyses for each trait-associated SNPs in genomic intervals separately. One integer indicates how many traits are analyzed.

Value

The data frame tb has columns,

<table>
<thead>
<tr>
<th>Trait</th>
<th>p.value</th>
<th>q.value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trait</td>
<td>Name of trait</td>
<td>P-value calculated from hypothesis testing</td>
</tr>
</tbody>
</table>
odds.ratio  Odds ratio calculated based on number of trait-associated SNPs in genomic intervals, number of trait-associated SNPs across whole genome, genomic intervals size (bps) and genome size (bps)
taSNP.hits  Number of trait-associated SNPs in genomic intervals
taSNP.num   Number of SNPs for specific trait

Author(s)
Li Chen <li.chen@emory.edu>, Zhaohui S.Qin<zhaohui.qin@emory.edu>

See Also
print.traseR

Examples
data(taSNPDB)
data(Tcell)
x=traseR(snpdb=taSNPDB,region=Tcell)
print(x)
Index

*Topic package
  traseR-package, 2

CEU, 2

plotContext (plots), 3
plotGene (plots), 3
plotInterval (plots), 3
plotPvalue (plots), 3
plots, 3
plotSNP (plots), 3
print.traseR, 4

queryGene (querys), 5
queryKeyword (querys), 5
querys, 5
querySNP (querys), 5

taSNPDB, 6
taSNPLDDB, 7
Tcell, 8
traseR, 8
traseR-package, 2