

# Package ‘traseR’

May 27, 2023

**Type** Package

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

**Version** 1.31.0

**Depends** R (>= 3.2.0), GenomicRanges, IRanges, BSgenome.Hsapiens.UCSC.hg19

**Suggests** BiocStyle, RUnit, BiocGenerics

**Date** 2021-11-21

**Author** Li Chen, Zhaohui S.Qin

**Maintainer** li.chen<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

**git\_url** <https://git.bioconductor.org/packages/traseR>

**git\_branch** devel

**git\_last\_commit** f95c5ca

**git\_last\_commit\_date** 2023-04-25

**Date/Publication** 2023-05-26

## R topics documented:

traseR-package	2
CEU	2
plots	3
print.traseR	4

quers	5
taSNP	6
taSNPLD	7
Tcell	8
traseR	8

<b>Index</b>	<b>11</b>
--------------	-----------

---

traseR-package	<i>GWAS trait-associated SNP enrichment analyses in genomic intervals</i>
----------------	---

---

### Description

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

### Details

Package: traseR  
 Type: Package  
 Version: 1.0  
 Date: 2015-11-18  
 License: GPL

### Author(s)

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

---

CEU	<i>Sampled SNPs from all SNPs of CEU population in 1000 genome project</i>
-----	--

---

### Description

A GRange 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>

---

plots *Visualize of trait-associated SNPs*

---

**Description**

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

**Usage**

```
plotContext(snpdb, region=NULL, keyword = NULL, pvalue = 1e-3)
plotPvalue(snpdb, region=NULL, keyword = NULL, 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**

plotContext	A pie plot with the distribution of SNP function class
plotPvalue	A density plot of -logPvalue of trait-associated SNPs
plotSNP	A plot of trait-associated SNP on chromosome
plotGene	A plot with the gene and possible nearby trait-associated SNPs
plotInterval	A plot with chromosome interval with possible nearby genes and trait-associated SNPs

**Author(s)**

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

**Examples**

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

---

print.traseR                      *Print the outcome of taSNP enrichment analyses*

---

**Description**

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

**Usage**

```
## S3 method for class 'traseR'
print(x, isTopK1=FALSE, topK1=10, isTopK2=FALSE, topK2=10, trait.threshold=10, traitclass.threshold=10, ..
```

**Arguments**

x	Object returned from traseR
isTopK1	If isTopK1 is TRUE, topK1 traits are printed; otherwise, traits with p-value below Bonferroni correction threshold are printed. Default is FALSE.
topK1	Top K1 traits are printed. Default is 10.
isTopK2	If isTopK2 is TRUE, topK2 trait class are printed; otherwise, trait class with p-value below Bonferroni correction threshold are printed. Default is FALSE.
topK2	Top K2 trait class are printed. Default is 10.
trait.threshold	Traits above this threshold are reported. Default is 10.
traitclass.threshold	Trait class 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**

```
data(taSNP)
data(Tcell)
x=traseR(snpdb=taSNP,region=Tcell)
print(x)
```

---

querys

*Retrieve trait-associated SNPs based*

---

**Description**

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

**Usage**

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

---

 taSNP

*trait-associated SNPs in dbGaP and NHGRI downloaded from Association Results Browser*

---

**Description**

A GRange object taSNP contains trait-associated SNPs from dbGaP and NHGRI downloaded from Association Results Browser.

**Usage**

```
data(taSNP)
```

**Value**

The data frame taSNP contains the following columns

Trait	Trait
Trait Class	Trait class which is formed based on the phenotype tree. Close traits are grouped together to form one class
SNP_ID	SNP rs number
p.value	GWAS SNP p-value
seqnames	Chromosome
ranges	Chromosome position
Context	SNP functional class
GENE_NAME	Nearest gene name
GENE_START	Gene start genomic position
GENE_END	Gene end genomic position
GENE_STRAND	Gene strand

**Author(s)**

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

---

taSNPLD	<i>linkage disequilibrium (&gt;0.8) within 100kb SNPs of all trait-associated SNPs from dbGaP and NHGRI</i>
---------	---

---

**Description**

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

**Usage**

```
data(taSNPLD)
```

**Value**

The data frame taSNPLD 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
Trait Class	Trait class which is formed based on the phenotype tree. Close traits are grouped together to form one class

**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, traitclass.threshold=0, pvalue = 1e-3)
```



**Arguments**

<code>snpdb</code>	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, <code>Source</code> , <code>Trait</code> , <code>SNP</code> , <code>p.value</code> , <code>Chr</code> , <code>Position</code> . 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, <code>SNP</code> , <code>Chr</code> , <code>Position</code> .
<code>region</code>	A GRange object or data frame, which is genomic intervals with three columns, chromosome, genomic start position, genomic end position.
<code>snpdb.bg</code>	A GRange object contains non-trait-associated SNPs. They are treated as background for statistical testing instead of whole genome as background if specified.
<code>keyword</code>	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.
<code>rankby</code>	Traits could be ranked by either p-value or <code>adds.ratio</code> based on the enrichment level of trait-associated SNPs in genomic intervals.
<code>test.method</code>	Several hypothesis testing options are provided: <code>binomial</code> (binomial test), <code>fisher</code> (Fisher's exact test), <code>chi sq</code> (Chi-squared test), <code>chi sq</code> (nonparametric test). Default is <code>binomial</code> (binomial test)
<code>alternative</code>	Indicate the alternative hypothesis. If <code>greater</code> , test if the genomic intervals are enriched in trait-associated SNPs than background. If <code>less</code> , test if the genomic intervals are depleted in trait-associated SNPs than background. If <code>two.sided</code> , test if there is difference between the enrichment of trait-associated SNPs in genomic intervals and in background.
<code>ntimes</code>	The number of shuffling time for one batch. See <code>nbatch</code> .
<code>nbatch</code>	The number of batches. The product of <code>ntimes</code> and <code>nbatch</code> is the total number of shuffling time.
<code>trait.threshold</code>	Test traits with number of SNPs more than the threshold.
<code>traitclass.threshold</code>	Test trait class with number of SNPs more than the threshold.
<code>pvalue</code>	SNPs with p-value less than this threshold are used for analyses.

**Details**

Return a list that contains three data frames. One data frame `tb.all` contains the results of enrichment analyses for all trait-associated SNPs in genomic intervals. Another data frame `tb1` contains the results of enrichment analyses for each trait-associated SNPs in genomic intervals separately. Another data frame `tb2` contains the results of enrichment analyses for each trait-class-associated SNPs in genomic intervals separately.

**Value**

The data frame `tb1` has columns,

<code>Trait</code>	Name of trait
--------------------	---------------

<code>p.value</code>	P-value calculated from hypothesis testing
<code>q.value</code>	Adjusted p-value from multiple testing using FDR correction
<code>odds.ratio</code>	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)
<code>taSNP.hits</code>	Number of trait-associated SNPs in genomic intervals
<code>taSNP.num</code>	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(taSNP)
data(Tcell)
x=traseR(snpdb=taSNP,region=Tcell)
print(x)
```

# Index

## \* 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

taSNP, 6

taSNPLD, 7

Tcell, 8

traseR, 8

traseR-package, 2