Package 'rxSeq'

December 1, 2016

Type Package

Title Combined Total and Allele Specific Reads Sequencing Study

Version 0.99.3

Date 2016-11-30

Author Vasyl Zhabotynsky [aut, cre], Wei Sun [aut], Fei Zou [aut]

Maintainer Vasyl Zhabotynsky <vasyl@unc.edu>

Description

Analysis of combined total and allele specific reads from the reciprocal cross study with RNA-seq data.

Depends MASS,VGAM,numDeriv

License GPL (>= 2)

LazyLoad yes

LazyData yes

Encoding UTF-8

NeedsCompilation no

Repository CRAN

Date/Publication 2016-12-01 12:45:14

R topics documented:

`data.A` *Sample data example for autosomal genes*

Description

This data set provides with example of experimental data for a subset of autosomal genes. The full model requires a combination of total read counts (**y**) - all the reads belonging for a gene, and finding out which of these reads we can specifically attribute to allele A or allele B - allele specific counts (**n**), separately the reads attributed specifically to allele B (**n0B**). Also, it includes the other data pieces to fit the model: **kappas** - total number of counts for each mouse, on log scale, index - specifying which cross each mouse belongs to, and geneids - Ensembl ids of genes. They, as well as the datasets simulated with **simRX** can be fitted using **proc.trecase.A** or **proc.trec.A**.

Value

`index`	vector defining the cross of the mouse, female - AB=1, BA=2, AA=3, BB=4, and male - AB=5, BA=6, AA=7, BB=8. If mice are of only one sex, AB=1, BA=2, AA=3, BB=4.
`y`	matrix of TReC counts. Note, the expected input assumes that inbred mice will be in the last columns of the table, after the last F1 mouse.
`n`	matrix of ASE counts for corresponding F1 mouse (classes 1, 2, 5, 6) for corresponding genes.
`n0B`	matrix of ASE counts belonging for allele B, for correponding genes and mice as in **n**.
`kappas`	A parameter, specifying log(overall TReC) for each mouse.
`geneids`	ids of genes, if not provided, rownames of the matrix **y** will be used

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

`process`,`readCounts`.

Examples

```
# see total read counts (TReC) for first 2  autosomal genes of a data example:
data.A$y[1:2,]
```

data.X *Sample data example for X chromosome genes*

Description

This data set provides with example of experimental data for a subset of autosomal genes. The full model requires a combination of total read counts (**y**) - all the reads belonging for a gene, and finding out which of these reads we can specifically attribute to allele A or allele B - allele specific counts (**n**), separately the reads attributed specifically to allele B (**n0B**). Also, **tausB** - is the *Xce* effect for each F1 mouse, which specifies the proportion of allele specific reads belonging to allele B. Also, it includes the other data pieces to fit the model: **kappas** - total number of counts for each mouse, on log scale, index - specifying which cross each mouse belongs to, and geneids - Ensembl ids of genes. They, as well as the datasets simulated with **simRX** can be fitted using **proc.trecase.X** or **proc.trec.X**.

Value

index	vector defining the cross of the mouse, female - AB=1, BA=2, AA=3, BB=4, and male - AB=5, BA=6, AA=7, BB=8. If mice are of only one sex, AB=1, BA=2, AA=3, BB=4.
y	matrix of TReC counts. Note, the expected input assumes that inbred mice will be in the last columns of the table, after the last F1 mouse.
n	matrix of ASE counts for corresponding F1 mouse (classes 1,2,5,6) for corresponding genes.
n0B	matrix of ASE counts belonging for allele B, for correponding genes and mice as in **n**.
kappas	A parameter, specifying as overall TReC for the mouse, on log scale
tausB	*Xce* effect: expression of allele B relative to the overall allele specific count for each mouse. Use some allele specific counts to estimate the effect.
geneids	ids of genes, if not provided, rownames of the matrix **y** will be used
index	index defining which mouse belongs to which cross
y	modified total read counts
n	modified allele specific counts
n0B	modified allele specific coutns, belonging to allele B
kappas	offset, defining a library size for each mouse
tausB	*Xce* effect for each mouse, for a given cross
geneids	Ensembl gene ids

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

process,readCounts.

Examples

```
# see total read counts (TReC) for first 2 X chromosome genes of a data example:
data.X$y[1:2,]
```

`get.tausB` *Produce* Xce *estimates for mice with allele specific reads*

Description

Xce estimation for mice with allele specific reads.

Usage

```
get.tausB(n, n0B, geneids, min.cnt=50, exclude.prop=.05, Xist.ID="ENSMUSG00000086503")
```

Arguments

`n`	vector of allele specific counts for each mouse
`n0B`	vector of allele specific counts for allele B
`geneids`	gene IDs
`min.cnt`	minimum number of allele specific counts
`exclude.prop`	minimum proportion of allele specific counts for each allele
`Xist.ID`	and ID of *Xist*, to exclude it from estimating *Xce*, since *Xce* would 1-tausB

Value

output - matrix of 4 rows:

`med.tauB`	taus estimated via median
`ave.tauB`	taus estimated via percent of allele B counts
`all.genes`	number of genes that had passed minimum count
`used.genes`	number of genes that had required percent of each allele

each column represent respective mouse.

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

process,nLogLik, data.X, rcX.

Examples

```
# Estimating XCE effect for each mouse for X chromosome
get.tausB(n=data.X$n, n0B=data.X$n0B, geneids=data.X$geneids)
```

---

nLogLik *Negative log likelihood for coefficients provided in results of the fit using process function*

Description

Calculates negative log(likelihood) of an X chromosome joint TReC and ASE counts model at a given set of parameters

Usage

```
nLogLik(res, rc, genei, hessian=FALSE)
```

Arguments

res	result object from process function
rc	Read count data object created by readCounts function
genei	get results for i'th gene
hessian	a logical option whether to calculate a Hessian matrix, the default values is set to FALSE.

Value

output - list(nll=-log.likelihood,hessian=hessian matrix)

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

process, rcA, readCounts.

Examples

```
## Not run:
# get negative-log likelihood at the given point:
nLogLik(res=trecase.X.out, rc=rcX, genei=1, hessian=TRUE)

## End(Not run)
```

process *Optimization wrapper, maximizing either the joint model of total (TReC) and allele specific (ASE) counts or just TReC*

Description

Performs optimization of one of four combinations: joint TReC and ASE or just TReC for autosome or X chromosome and tests with lrt test several hypotheses: additive, parent of origin, dominance, consistency of TreC and ASE additive effect, ASE only additive effect, sex, sex specific additive, dominance deviation for males.

Usage

```
process(rc, hessian=FALSE)
```

Arguments

rc	an object of class readCounts.
hessian	a flag whether Hessian matrix for these genes should be calculated, by default set to FALSE

Value

a list of following matrices (if there is only one sex, only the relevant tests and matrices are outputed) :

pvals	matrix of p-values from description for each gene corresponding row
coef.full	matrix of full model fit coefficients, -log(likelihood at these coefficients),phi, theta (2 overdispersion parameters used)
coef.add	matrix of additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
coef.poo	matrix of parent of origin restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
coef.dom	matrix of dominance restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
coef.same	matrix of TReC=ASE additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
coef.ase.add	matrix of ASE additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
coef.sex	matrix of sex restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
coef.sex.add	matrix of sex specific additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
coef.dev.dom	matrix of dominance deviation for male restricted fit coefficients, -log(likelihood at these coefficients),phi, theta

errorlist	a list of errors
hess.lst	a list of heassian matrices, if parameter **hessian** is set to TRUE

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

get.tausB,nLogLik, data.X, data.A, rcA, readCounts.

Examples

```
## Not run:
# fitting X chromosome data example, for a full model, i.e. assuming we have allele specific reads:
trecase.A.out = process(rc=rcA)
names(trecase.A.out)
trecase.A.out$pval

#alternatively for X chromosome:
trecase.X.out = process(rc=rcX)
names(trecase.X.out)
trecase.X.out$pval

## End(Not run)
```

---

rcA *Reformatted data for autosomal set to be used as input to process function*

Description

This is an object of type readCounts provides with example of experimental data for a subset of autosomal genes. The full model requires a combination of total read counts (**y**) - all the reads belonging for a gene, and finding out which of these reads we can specifically attribute to allele A or allele B - allele specific counts (**n**), separately the reads attributed specifically to allele B (**n0B**). In autosomes *Xce* effect is absent, so it would be set to NULL for this dataset. Also, it includes the other data pieces to fit the model: **kappas** - total number of counts for each mouse, on log scale, index - specifying which cross each mouse belongs to, and geneids - Ensembl ids of genes. Such values also can be simulated with **simRX** can be fitted using **process** with appropriate options chrom="X" and field model to be either "full" or "short".

Value

index	vector defining the cross of the mouse, female - AB=1, BA=2, AA=3, BB=4, and male - AB=5, BA=6, AA=7, BB=8. If mice are of only one sex, AB=1, BA=2, AA=3, BB=4.

y	matrix of TReC counts. Note, the expected input assumes that inbred mice will be in the last columns of the table, after the last F1 mouse.
n	matrix of ASE counts for corresponding F1 mouse (classes 1,2,5,6) for corresponding genes.
n0B	matrix of ASE counts belonging for allele B, for correponding genes and mice as in **n**.
kappas	A parameter, specifying as overall TReC for the mouse, on log scale
tausB	*Xce* effect: expression of allele B relative to the overall allele specific count for each mouse. Set to NULL in autosomes.
gene.switch	For which genes *Xce* effect should be switched. Null for autosomes.
geneids	ids of genes, if not provided, rownames of the matrix **y** will be used
chrom	this field would be set to be "X" since this is dataset for chromosome X
model	set to be "full", can be modified to "short" to run a TReC oply model
geneids	Ensembl gene ids
tech.ctrl	a list of overdispersion boundaries and log(2)

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

process,readCounts, simRX.

Examples

```
# see total read counts (TReC) for first 2 X chromosome genes of a data example:
rcX
```

---

rcX *Reformatted data for chromosome X set to be used as input to process function*

Description

This is an object of type readCounts provides with example of experimental data for a subset of X chromosome genes. The full model requires a combination of total read counts (**y**) - all the reads belonging for a gene, and finding out which of these reads we can specifically attribute to allele A or allele B - allele specific counts (**n**), separately the reads attributed specifically to allele B (**n0B**). Also, **tausB** - is the *Xce* effect for each F1 mouse, which specifies the proportion of allele specific reads belonging to allele B. Also, it includes the other data pieces to fit the model: **kappas** - total number of counts for each mouse, on log scale, index - specifying which cross each mouse belongs to, and geneids - Ensembl ids of genes. They, as well as the datasets simulated with **simRX** can be fitted using **process** with appropriate options chrom="X" and field model to be either "full" or "short".

Value

genes.switch=genes.switch, geneids=geneids,

index	vector defining the cross of the mouse, female - AB=1, BA=2, AA=3, BB=4, and male - AB=5, BA=6, AA=7, BB=8. If mice are of only one sex, AB=1, BA=2, AA=3, BB=4.
y	matrix of TReC counts. Note, the expected input assumes that inbred mice will be in the last columns of the table, after the last F1 mouse.
n	matrix of ASE counts for corresponding F1 mouse (classes 1,2,5,6) for corresponding genes.
n0B	matrix of ASE counts belonging for allele B, for correponding genes and mice as in **n**.
kappas	A parameter, specifying as overall TReC for the mouse, on log scale
tausB	*Xce* effect: expression of allele B relative to the overall allele specific count for each mouse. Use some allele specific counts to estimate the effect.
gene.switch	For which genes *Xce* effect should be switched. Xist gene set to be switched in this set.
geneids	ids of genes, if not provided, rownames of the matrix **y** will be used
chrom	this field would be set to be "X" since this is dataset for chromosome X
model	set to be "full", can be modified to "short" to run a TReC oply model
geneids	Ensembl gene ids
tech.ctrl	a list of overdispersion boundaries and log(2)

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

process,readCounts,simRX.

Examples

```
# see total read counts (TReC) for first 2 X chromosome genes of a data example:
rcX
```

`readCounts` *A list object that should be used as input to optimization process function.*

Description

It should contain at least total read counts (TReC) and classification of crosses 1 to 8. To fit the full model should also have appropriate allele specific counts n and n0B. Also is used along with results of optimization as input to nLogLik function if one needs to calculate Hessian matrix.

Value

`index`	vector defining the cross of the mouse, female - AB=1, BA=2, AA=3, BB=4, and male - AB=5, BA=6, AA=7, BB=8. If mice are of only one sex, AB=1, BA=2, AA=3, BB=4.
`y`	matrix of TReC counts. Note, the expected input assumes that inbred mice will be in the last columns of the table, after the last F1 mouse.
`n`	matrix of ASE counts for corresponding F1 mouse (classes 1,2,5,6) for corresponding genes.
`n0B`	matrix of ASE counts belonging for allele B, for correponding genes and mice as in **n**.
`kappas`	A parameter, specifying as overall TReC for the mouse, on log scale
`tausB`	*Xce* effect: expression of allele B relative to the overall allele specific count for each mouse. Set to NULL in autosomes.
`gene.switch`	For which genes *Xce* effect should be switched. Null for autosomes.
`geneids`	ids of genes, if not provided, rownames of the matrix **y** will be used
`chrom`	this field would be set to be "X" since this is dataset for chromosome X
`model`	set to be "full", can be modified to "short" to run a TReC oply model
`geneids`	Ensembl gene ids
`tech.ctrl`	a list of overdispersion boundaries and log(2)

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

`process`,`nLogLik`, `simRX`.

Examples

```
# see total read counts (TReC) for first 2 X chromosome genes of a data example:
rcA = readCounts(index=data.A$index, y=data.A$y[1:2,], n=data.A$n[1:2,], n0B=data.A$n0B[1:2,],
                 kappas=data.A$kappas, geneids=data.A$geneids[1:2])
```

simRX *Produce simulated counts*

Description

This function is producing simulated counts for the joint model with Negative-Binomial distribution for TReC and Beta-Binomial for ASE counts. The simulated dataset should be reformatted to readCounts format to be used for optimization.

Usage

```
simRX(b0f, b0m, b1f, b1m, beta_sex, beta_dom, beta_k=1, phi=1, theta=1, n=6,
      mean.base.cnt=50, range.base.cnt=60, perc.ase=.35, n.simu=1E4,
      is.X=FALSE, tauB=NULL, seed=NULL)
```

Arguments

b0f	a female additive strain effect
b0m	a male additive strain effect
b1f	a female parent of origin effect
b1m	a male parent of origin effect
beta_sex	a sex effect
beta_dom	a dominance effect
beta_k	an effect associated with the library size kappas
phi	a Negative-Binomial overdispersion, default value is 1
theta	a Beta-Binomial overdispersion, default value is 1
n	a vector defining number of mice in each cross, default value is 6
mean.base.cnt	a target expected number of counts for the base group (with no effects), default value is 50
range.base.cnt	a range in which the expected number of counts for the base group will vary, default value is 60
perc.ase	a percent reads that are allele-specific, default value is 35%
n.simu	a number of simulations, default value is 1E4
is.X	a flag if the value to be simulated is X for chromosome (otherwise autosome), default value is FALSE
tauB	a value describing allelic imbalance - *Xce* effect for the cross, default value is NULL, in which case 50% will be simulated
seed	a random seed to be set, no set by default.

Value

output - 3 matrices with one row - one gene, one column - one mouse:

index	vector defining the cross of the mouse, female - AB=1, BA=2, AA=3, BB=4, and male - AB=5, BA=6, AA=7, BB=8. If mice are of only one sex, AB=1, BA=2, AA=3, BB=4.
y	A matrix of total read counts
n	A matrix of allele specific counts
n0B	A matrix of allele specific counts associated with allele B
kappas	Offset parameter, given as overall TReC for the mouse.
tausB	In case of the simulating X chromosome the provided *Xce* effect is returned: expression of allele B relative to the overall allele specific count for each mouse.

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

process,readCounts.

Examples

```
# simulating autosomal data:
dat.A = simRX(b0f=.5, b0m=.6, b1f=.3, b1m=.4, beta_sex=.1, beta_dom=.1, n.simu=1E1)
names(dat.A)
# simulating autosomal data:
dat.X = simRX(b0f=.5, b0m=.6, b1f=.3, b1m=.4, beta_sex=.1, beta_dom=.1, n.simu=1E1,
              is.X=TRUE, tauB=.3)
names(dat.X)
```

---

trecase.A.out *Example of results produced by optimizing step using process function on autosomal genes. Structured as a list.*

Description

A list containing test results as well as parameter estimates for joint model evaluated by process function for autosomal genes.

Value

a list of following matrices (if there is only one sex, only the relevant tests and matrices are outputed) :

`pvals`	matrix of p-values from description for each gene corresponding row
`coef.full`	matrix of full model fit coefficients, -log(likelihood at these coefficients),phi, theta (2 overdispersion parameters used)
`coef.add`	matrix of additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.poo`	matrix of parent of origin restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.dom`	matrix of dominance restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.same`	matrix of TReC=ASE additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.ase.add`	matrix of ASE additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.sex`	matrix of sex restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.sex.add`	matrix of sex specific additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.dev.dom`	matrix of dominance deviation for male restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`errorlist`	a list of errors
`hess.lst`	a list of heassian matrices, if parameter **hessian** is set to TRUE

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

`process`, `rcA`, `data.A`.

Examples

```
names(trecase.A.out)
```

`trecase.X.out` *Example of results produced by optimizing step using process function on X chromosome genes. Structured as a list.*

Description

A list containing test results as well as parameter estimates for joint model evaluated by process function for autosomal genes.

Value

a list of following matrices (if there is only one sex, only the relevant tests and matrices are outputed) :

`pvals`	matrix of p-values from description for each gene corresponding row
`coef.full`	matrix of full model fit coefficients, -log(likelihood at these coefficients),phi, theta (2 overdispersion parameters used)
`coef.add`	matrix of additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.poo`	matrix of parent of origin restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.dom`	matrix of dominance restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.same`	matrix of TReC=ASE additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.ase.add`	matrix of ASE additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.sex`	matrix of sex restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.sex.add`	matrix of sex specific additive restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`coef.dev.dom`	matrix of dominance deviation for male restricted fit coefficients, -log(likelihood at these coefficients),phi, theta
`errorlist`	a list of errors
`hess.lst`	a list of heassian matrices, if parameter **hessian** is set to TRUE

Author(s)

Vasyl Zhabotynsky vasyl@unc.edu

See Also

`process`, `rcX`, `data.X`.

Examples

```
names(trecase.X.out)
```

Index