

# Package ‘deepSNV’

September 24, 2012

**Imports** Rsamtools

**Maintainer** Moritz Gerstung <gemoritz@ethz.ch>

**License** GPL-3

**Title** Test for subclonal SNVs in deep sequencing experiments.

**BiocViews** GeneticVariability, SNP, Sequencing, Genetics, DataImport

**LinkingTo** Rsamtools

**Type** Package

**LazyLoad** yes

**Author** Moritz Gerstung and Niko Beerenwinkel

**Description** This package provides a quantitative variant caller for detecting subclonal mutations in ultra-deep ( $\geq 100x$  coverage) sequencing experiments. It assumes a comparative setup with a control experiment of the same loci and a beta-binomial model to discriminate sequencing errors and subclonal SNVs.

**Version** 1.2.3

**URL** <http://www.cbg.ethz.ch/software/deepSNV>

**Depends** R ( $\geq 2.13.0$ ), Rsamtools ( $\geq 1.4.3$ ), GenomicRanges, IRanges, Biostrings, VGAM, methods, graphics

**Collate** ‘deepSNV-class.R’ ‘deepSNV-experimental.R’ ‘deepSNV-functions.R’ ‘deepSNV-generics.R’ ‘deepSNV-methods.R’ ‘deepSNV-misc.R’ ‘deepSNV-package.R’

## R topics documented:

deepSNV-package . . . . .	2
bam2R . . . . .	3
consensusSequence-methods . . . . .	4
control-methods . . . . .	5
coordinates-methods . . . . .	6
deepSNV-class . . . . .	6
deepSNV-methods . . . . .	8
estimateDirichlet-methods . . . . .	10

estimateDispersion-methods . . . . .	10
Extract-methods . . . . .	11
manhattanPlot . . . . .	12
normalize-methods . . . . .	13
p.combine . . . . .	14
p.val-methods . . . . .	15
phiX . . . . .	15
plot.deepSNV . . . . .	16
RCC . . . . .	17
repeatMask-methods . . . . .	17
RF . . . . .	18
show,deepSNV-method . . . . .	19
summary-methods . . . . .	20
test-methods . . . . .	21
trueSNVs . . . . .	22

## Index 23

---

deepSNV-package	<i>Estimation of single nucleotide variants from paired deep sequencing experiments</i>
-----------------	---

---

## Description

Estimation of single nucleotide variants from paired deep sequencing experiments

## Details

Package:	deepSNV
Type:	Package
Version:	0.99.3
Date:	2012-01-20
License:	GPL3
LazyLoad:	yes

This packages provides algorithms for estimating SNVs and their frequencies from matched ultra-deep sequencing experiments of heterogeneous samples. It retrieves the nucleotide counts at each position and each strand and tests for differences between the two experiments with a likelihood ratio test using either a binomial or beta-binomial model. The statistic can be tuned across genomic sites by a shared Dirichlet prior.

## Author(s)

Moritz Gerstung, ETH Zurich, D-BSSE <gemoritz@ethz.ch>

## References

See our upcoming paper ...

**See Also**[deepSNV](#)**Examples**

```
## Short example with 2 SNVs at frequency ~10%
regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIIB_BRU_K034", start = 3120, stop=3140)
ex <- deepSNV(test = system.file("extdata", "test.bam", package="deepSNV"), control = system.file("extdata", "control.bam", package="deepSNV"))
show(ex) # show method
plot(ex) # scatter plot
summary(ex) # summary with significant SNVs
ex[1:3,] # subsetting the first three genomic positions
tail(test(ex, total=TRUE)) # retrieve the test counts on both strands
tail(control(ex, total=TRUE))

## Not run: Full example with ~ 100 SNVs. Requires an internet connection, but try yourself.
# regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIIB_BRU_K034", start = 2074, stop=3585)
# HIVmix <- deepSNV(test = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/test.bam", control = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/control.bam")
data(HIVmix) # attach data instead.
show(HIVmix)
plot(HIVmix)
head(summary(HIVmix))
```

bam2R

*Read nucleotide counts...***Description**

Read nucleotide counts

**Usage**

```
bam2R(file, chr, start, stop, q=25, s=2, head.clip=0, max.depth=1e+06)
```

**Arguments**

file	The name of the .bam file as a string.
chr	The chromosome as a string.
start	The start position (1-indexed).
stop	The end position (1-indexed).
q	An optional cutoff for the nucleotide Phred quality. Default q = 25. Nucleotides with Q < q will be masked by 'N'.
s	Optional choice of the strand. Defaults to s = 2 (both).
head.clip	Should n nucleotides from the head of reads be clipped? Default 0.
max.depth	The maximal depth for the pileup command. Default 1,000,000.

**Details**

This function reads the nucleotide counts on each position of a .bam alignment. The counts of both strands are reported separately and nucleotides below a quality cutoff are masked. It is called by [deepSNV](#) to parse the alignments of the test and control experiments, respectively.

**Value**

A named `matrix` with rows corresponding to genomic positions and columns for the nucleotide counts (A, T, C, G, -), masked nucleotides (N), (INS)ertions, (DEL)etions, (HEAD)s and (TAIL)s that count how often a read begins and end at the given position, respectively, and the sum of alignment (QUAL)ities. Counts from matches on the reference strand (s=0) are uppercase, counts on the complement (s=1) are lowercase. The returned matrix has  $11 * 2$  (strands) = 22 columns and (stop - start + 1) rows.

**Author(s)**

gemoritz

**Examples**

```
## Simple example:
counts <- bam2R(file = system.file("extdata", "test.bam", package="deepSNV"), chr="B.FR.83.HXB2_LAI_IIIB_B
show(counts)
## Not run: Requires an internet connection, but try yourself.
# bam <- bam2R(file = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/test.bam", chr="B.FR.83.HXB2_LAI_
# head(bam)
```

---

consensusSequence-methods

*Calculate the consensus sequence.*

---

**Description**

Calculate the consensus sequence.

**Usage**

```
## S4 method for signature 'matrix'
consensusSequence(x, vector=FALSE, haploid=TRUE, het.cut=0.333)
## S4 method for signature 'deepSNV'
consensusSequence(x, vector=FALSE, haploid=TRUE, het.cut=0.333)
```

**Arguments**

<code>x</code>	An object. Either an <code>deepSNV-class</code> object, or a named matrix with nucleotide counts.
<code>vector</code>	Boolean where TRUE indicates that a character vector should be returned.
<code>haploid</code>	Should the consensus be called for a haploid control? Otherwise, also all bases larger than <code>het.cut</code> are reported. Default <code>haploid = TRUE</code> .
<code>het.cut</code>	Heterozygous cutoff. If <code>haploid = FALSE</code> , report all nucleotides with relative frequency larger than <code>het.cut</code> . Default = 0.333.

**Details**

`consensusSequence-methods`: This function computes the consensus sequence from a matrix of nucleotide counts, or the control slot of a `deepSNV` object.

**Value**

consensusSequence-methods: A [DNASTring](#) with the consensus sequence, or if vector = TRUE, a character vector.

**Author(s)**

gemoritz

**Examples**

```
data(HIVmix)
seq = consensusSequence(HIVmix)
consensusSequence(HIVmix, vector=TRUE)[1:10]
```

---

control-methods

*Get control counts...*

---

**Description**

Get control counts

**Usage**

```
## S4 method for signature 'deepSNV'
control(deepSNV, total=F)
```

**Arguments**

deepSNV            a [deepSNV-class](#) object  
total              Logical. If true the sum of both strands is returned

**Details**

control-methods: Convenience function to obtain the control counts from a deepSNV object.

**Value**

control-methods: A matrix with the absolute frequencies summed over both strands.

**Examples**

```
data(HIVmix)
control(HIVmix)[1:10,]
control(HIVmix, total=TRUE)[1:10,]
```

---

coordinates-methods    *Get coordinates...*

---

### Description

Get coordinates

### Usage

```
## S4 method for signature 'deepSNV'
coordinates(deepSNV)
```

### Arguments

deepSNV            a [deepSNV-class](#) object

### Details

coordinates-methods: Convenience function to get the coordinates from a deepSNV object.

### Value

coordinates-methods: A [data.frame](#) with columns "chrom(osome)" and "pos(ition)".

### Examples

```
data(HIVmix)
coordinates(HIVmix)[1:10,]
```

---

deepSNV-class            *deepSNV class.*

---

### Description

deepSNV class.

### Details

This class stores the contents of the deepSNV test. It is typically initialized with [deepSNV-methods](#).

### Slots

p.val: ([matrix](#)) The P-values of the test.

test: ([matrix](#)) A matrix with the nucleotide counts in the test experiment. The column names of the nucleotide counts are A, T, C, G, - for the positive strand and a, t, c, g, \_ for the reverse.

control: ([matrix](#)) A matrix with the nucleotide counts in the control experiment. The column names must be the same as for the test.

coordinates: ([data.frame](#)) A [data.frame](#) with the genomic coordinates chr and pos, and other columns, if desired.

**dirichlet.prior:** ([matrix](#)) A matrix with the nucleotide-specific Dirichlet prior  
**alternative:** ([character](#)) A string with the alternative used in the test.  
**nucleotides:** ([vector](#)) A character vector with the nucleotides tested.  
**regions:** ([data.frame](#)) A [data.frame](#) with columns chr, start, and stop.  
**files:** ([list](#)) A list with two entries test and control storing the filenames (if the object was initialized from two bam-files).  
**combine.method:** ([character](#)) The method for combining p-values as a character string.  
**model:** ([character](#)) The statistical model, either bin for binomial, or betabin for beta-binomial  
**over.dispersion:** ([numeric](#)) If the model is beta-binomial, the first parameter for the beta-binomial model, which is shared across sites.  
**call:** ([call](#)) The last function call to deepSNV.  
**log.lik:** ([numeric](#)) The log likelihood of the data under the null hypothesis. (Excluding zeros on the opposite site under a one-sided test.)

## Methods

[summary](#) signature(object = "deepSNV"): Summary for deepSNV object  
[\[](#) signature(x = "deepSNV"): Subsetting for deepSNV objects.  
[show](#) signature(object = "deepSNV"): Show method for deepSNV objects  
[deepSNV](#) signature(test = "deepSNV", control = "missing"): Method for signature deep-SNV,missing  
[estimateDirichlet](#) signature(control = "deepSNV"): Method for signature deepSNV  
[repeatMask](#)  
[consensusSequence](#)  
[estimateDispersion](#)  
[normalize](#) signature(test = "deepSNV", control = "missing"): Normalize nucleotide counts.  
[control](#) signature(deepSNV = "deepSNV"): Get controls counts.  
[test](#) signature(deepSNV = "deepSNV"): Get test counts.  
[p.val](#) signature(deepSNV = "deepSNV"): Get p-values  
[coordinates](#) signature(deepSNV = "deepSNV"): Get coordinates

## Author(s)

gemoritz

## See Also

[deepSNV-methods](#)

## Examples

```

## Short example with 2 SNVs at frequency ~10%
regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIB_BRU_K034", start = 3120, stop=3140)
ex <- deepSNV(test = system.file("extdata", "test.bam", package="deepSNV"), control = system.file("extdata", "control.bam", package="deepSNV"))
show(ex) # show method
plot(ex) # scatter plot
summary(ex) # summary with significant SNVs
  
```

```

ex[1:3,] # subsetting the first three genomic positions
tail(test(ex, total=TRUE)) # retrieve the test counts on both strands
tail(control(ex, total=TRUE))

## Not run: Full example with ~ 100 SNVs. Requires an internet connection, but try yourself.
# regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIB_BRU_K034", start = 2074, stop=3585)
# HIVmix <- deepSNV(test = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/test.bam", control = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/control.bam")
data(HIVmix) # attach data instead..
show(HIVmix)
plot(HIVmix)
head(summary(HIVmix))

```

---

deepSNV-methods	<i>Test two matched deep sequencing experiments for low-frequency SNVs.</i>
-----------------	---

---

## Description

Test two matched deep sequencing experiments for low-frequency SNVs.

## Usage

```

## S4 method for signature 'matrix,matrix'
deepSNV(test, control, alternative=c("greater", "less", "two.sided"),
        dirichlet.prior, pseudo.count=1, combine.method=c("fisher", "max",
        "average"), over.dispersion=100, model=c("bin", "betabin"), ...)
## S4 method for signature 'deepSNV,missing'
deepSNV(test, control, ...)
## S4 method for signature 'character,character'
deepSNV(test, control, regions, q=25, s=2, head.clip=0, ...)
## S4 method for signature 'matrix,character'
deepSNV(test, control, regions, q=25, s=2, ...)
## S4 method for signature 'character,matrix'
deepSNV(test, control, regions, q=25, s=2, ...)

```

## Arguments

test	The test experiment. Either a .bam file, or a matrix with nucleotide counts, or a <a href="#">deepSNV-class</a> object.
control	The control experiment. Must be of the same type as test, or missing if test is a <a href="#">deepSNV-class</a> object.
alternative	The alternative to be tested. One of greater, less, or two.sided.
model	Which model to use. Either "bin", or "betabin". Default "bin".
dirichlet.prior	A base-sepecific Dirichlet prior specified as a matrix. Default NULL.
pseudo.count	If dirichlet.prior=NULL, a pseudocount can be used to define a flat prior.
over.dispersion	A numeric factor for the over.dispersion, if the model is beta-binomial. Default 100.
combine.method	The method to combine p-values. One of "fisher" (default), "max", or "average". See <a href="#">p.combine</a> for details.



regions	The regions to be parsed if test and control are .bam files. Either a <a href="#">data.frame</a> with columns "chr" (chromosome), "start", "stop", or a <a href="#">GRanges</a> object. If multiple regions are specified, the appropriate slots of the returned object are concatenated by row.
q	The quality argument passed to <a href="#">bam2R</a> if the experiments are .bam files.
s	The strand argument passed to <a href="#">bam2R</a> if the experiments are .bam files.
head.clip	The head.clip argument passed to <a href="#">bam2R</a> if the experiments are .bam files.
...	Additional arguments.

## Details

deepSNV-methods: This generic function can handle different types of inputs for the test and control experiments. It either reads from two .bam files, uses two matrices of nucleotide counts, or re-evaluates the test results from a [deepSNV-class](#) object. The actual test is a likelihood ratio test of a (beta-)binomial model for the individual nucleotide counts on each position under the hypothesis that both experiments share the same parameter, and the alternative that the parameters differ. Because the difference in degrees of freedom is 1, the test statistic  $D = -2 \log \max L_0 / \max L_1$  is asymptotically distributed as  $\chi_1^2$ . The statistic may be tuned by a nucleotide specific Dirichlet prior that is learned across all genomic sites, see [estimateDirichlet](#). If the model is beta-binomial, a global dispersion parameter is used for all sites. It can be learned with [estimateDispersion](#).

## Value

deepSNV-methods: A [deepSNV](#) object

## Author(s)

gemoritz

## Examples

```
## Short example with 2 SNVs at frequency ~10%
regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIIB_BRU_K034", start = 3120, stop=3140)
ex <- deepSNV(test = system.file("extdata", "test.bam", package="deepSNV"), control = system.file("extdata", "control.bam", package="deepSNV"))
show(ex) # show method
plot(ex) # scatter plot
summary(ex) # summary with significant SNVs
ex[1:3,] # subsetting the first three genomic positions
tail(test(ex, total=TRUE)) # retrieve the test counts on both strands
tail(control(ex, total=TRUE))

## Not run: Full example with ~ 100 SNVs. Requires an internet connection, but try yourself.
# regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIIB_BRU_K034", start = 2074, stop=3585)
# HIVmix <- deepSNV(test = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/test.bam", control = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/control.bam")
data(HIVmix) # attach data instead..
show(HIVmix)
plot(HIVmix)
head(summary(HIVmix))
```

---

estimateDirichlet-methods

*Learn a base-specific Dirichlet prior.*

---

### Description

Learn a base-specific Dirichlet prior.

### Usage

```
## S4 method for signature 'matrix'  
estimateDirichlet(control)  
## S4 method for signature 'deepSNV'  
estimateDirichlet(control)
```

### Arguments

control            Either a matrix with nucleotide counts or a [deepSNV-class](#) object.

### Details

estimateDirichlet-methods: The prior learns the parameters of a Dirichlet distribution separately for each consensus base. The expected value of the Dirichlet distributions is the base-substitution matrix, where rows correspond to the initial nucleotide and columns to the substituted nucleotide. The absolute values determine the higher moments of the Dirichlet distributions. After having learned the prior the [deepSNV-class](#) test is recomputed.

### Value

estimateDirichlet-methods: An [deepSNV-class](#) object.

### Author(s)

gemoritz

### Examples

```
data(phiX)  
estimateDirichlet(phiX)
```

---

estimateDispersion-methods

*Estimate the Dispersion factor in a beta-binomial model.*

---

### Description

Estimate the Dispersion factor in a beta-binomial model.

**Usage**

```
## S4 method for signature 'deepSNV,missing'
estimateDispersion(test, control, alternative, interval=c(0, 1000))
## S4 method for signature 'matrix,matrix'
estimateDispersion(test, control, alternative, interval=c(0, 1000))
```

**Arguments**

test	Either a deepSNV object, or a matrix with the test counts.
control	Missing if test is a deepSNV object, otherwise missing.
alternative	The alternative to be tested. One of "greater", "less", "two-sided" (default). If test is a deepSNV object, automatically taken from the corresponding slot if unspecified.
interval	The interval to be screened for the overdispersion factor. Default (0,1000).

**Details**

estimateDispersion-methods: This function estimates the dispersion factor in a beta-binomial model of the nucleotide counts. This model assumes that the count for nucleotide  $j$  at position  $i$  is distributed after a beta-binomial  $X_{i,j} \sim \text{BB}(n_i; \alpha, \beta_{ij})$ , where  $n_i$  is the coverage. The base and nucleotide specific parameter  $\beta_{ij}$  is estimated from the local mean by the method-of-moments estimate,  $\alpha$  is a shared overdispersion parameter. It is estimated via a numerical optimization of the likelihood under the null-hypothesis.

**Value**

estimateDispersion-methods: A `deepSNV-class` object if the input was a deepSNV object. Otherwise the loglikelihood and the estimated parameter.

**Author(s)**

gemoritz

**Examples**

```
data("RCC", package="deepSNV")
plot(RCC)
summary(RCC)[,1:6]
RCC.bb = estimateDispersion(RCC, alternative = "two.sided")
summary(RCC.bb)
```

---

Extract-methods

*Subsetting for deepSNV objects.*

---

**Description**

Subsetting for deepSNV objects.

**Usage**

```
## S4 method for signature 'deepSNV'
x[i, j]
```

**Arguments**

x            A `deepSNV-class` object.  
i            Row indices.  
j            Column (nucleotide) indices.

**Value**

A `deepSNV-class` object.

**Author(s)**

gemoritz

**Examples**

```
data(HIVmix)
HIVmix[1:10,]
```

---

manhattanPlot

*Manhattan plot.*

---

**Description**

Manhattan plot.

**Usage**

```
manhattanPlot(x, col=nt.col)
```

**Arguments**

x            An `deepSNV` object.  
col          An optional vector of colors for the nucleotides.

**Details**

This functions performs a Manhattan plot of the p-values of a deepSNV test against the position

**Value**

NULL.

**Author(s)**

gemoritz

**Examples**

```
data(HIVmix)
manhattanPlot(HIVmix)
```

---

normalize-methods	<i>Normalize nucleotide counts.</i>
-------------------	-------------------------------------

---

## Description

Normalize nucleotide counts.

## Usage

```
## S4 method for signature 'matrix,matrix'  
normalize(test, control, round=TRUE, ...)  
## S4 method for signature 'deepSNV,missing'  
normalize(test, control, ...)
```

## Arguments

test	Either an <a href="#">deepSNV-class</a> object or a named matrix with nucleotide counts.
control	Missing if test is an <code>link{deepSNV-class}</code> object, otherwise a matrix with nucleotide counts.
round	Logical. Should normalized counts be round to integers? Default=TRUE
...	Parameters passed to <a href="#">loess</a> .

## Details

normalize-methods: This functions performs a [loess](#) normalization of the nucleotide. This experimental feature can be used to compare experiments from different libraries or sequencing runs that may have differing noise characteristics.

## Value

normalize-methods: A [deepSNV-class](#) object.

## Author(s)

gemoritz

## Examples

```
data(phiX, package = "deepSNV")  
plot(phiX)  
phiN <- normalize(phiX, round = TRUE)  
plot(phiN)
```

---

p.combine	<i>Combine two p-values...</i>
-----------	--------------------------------

---

### Description

Combine two p-values

### Usage

```
p.combine(p1, p2, method=c("fisher", "max", "average", "prod"))
```

### Arguments

p1	P-value 1
p2	P-value 2
method	One of "fisher" (default), "max" or "average"

### Details

This function combines two P-values into a single one using a statistic defined by method. "fisher" uses the product of the two, in this case the logarithm of the product is  $\chi_4^2$  distributed. If the method = "max", the resulting P-value is  $\max\{P_1, P_2\}^2$ . For method = "average" the mean is used, yielding a P-value of  $2x^2$  if  $x = (P_1 + P_2)/2 < .5$  and  $1 - 2x^2$  otherwise.

### Value

p-values

### Author(s)

gemoritz

### Examples

```
p1 <- runif(1000)
p2 <- runif(1000)
hist(p1)
p.avg = p.combine(p1,p2, method="average")
hist(p.avg)
p.fish = p.combine(p1,p2, method="fisher")
hist(p.fish)
p.max = p.combine(p1,p2, method="max")
hist(p.max)
pairs(data.frame(p1,p2,p.fish,p.max,p.avg))
```

---

p.val-methods	<i>Get p-values...</i>
---------------	------------------------

---

**Description**

Get p-values

**Usage**

```
## S4 method for signature 'deepSNV'  
p.val(deepSNV)
```

**Arguments**

deepSNV            a [deepSNV-class](#) object

**Details**

p.val-methods: Convenience function to get the p-values from a deepSNV object.

**Value**

p.val-methods: A matrix with the p-values.

**Examples**

```
data(HIVmix)  
p.val(HIVmix)[1:10,]
```

---

phiX	<i>Example phiX data...</i>
------	-----------------------------

---

**Description**

Example phiX data

**Details**

Data from two phiX experiments sequenced on a GAIIx.

**Examples**

```
data(phiX, package="deepSNV")  
plot(phiX)  
phiN <- normalize(phiX, round=TRUE)  
plot(phiN)
```

---

plot.deepSNV                      *Scatter plot of relative nucleotide frequencies.*

---

### Description

Scatter plot of relative nucleotide frequencies.

### Usage

```
## S3 method for class 'deepSNV'
plot(x, sig.level = NULL, col = NULL, col.null = "grey", cex.min = 0.2, ylab="Relative Frequency")
```

### Arguments

x	A deep SNV object.
sig.level	By default, p-values below sig.level are drawn as filled circles.
col	Color of the nucleotides.
col.null	Color of insignificant nucleotides.
cex.min	The minimal size of the points.
xlab	The x-axis label.
ylab	The y-axis label.
pch	The plotting symbol. Default = 16 (filled circle)
...	Additional arguments passed to plot.

### Details

This function plots the relative nucleotide frequencies of the test against the control experiment on a logarithmic scale. The color of the symbols denotes the nucleotide, and the area of the circle is proportional to the  $-\log$  of the p-value.

### Value

NULL

### Author(s)

gemoritz

### Examples

```
## Short example with 2 SNVs at frequency ~10%
regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIIB_BRU_K034", start = 3120, stop=3140)
ex <- deepSNV(test = system.file("extdata", "test.bam", package="deepSNV"), control = system.file("extdata", "control.bam", package="deepSNV"))
show(ex) # show method
plot(ex) # scatter plot
summary(ex) # summary with significant SNVs
ex[1:3,] # subsetting the first three genomic positions
tail(test(ex, total=TRUE)) # retrieve the test counts on both strands
tail(control(ex, total=TRUE))
```



```
## Not run: Full example with ~ 100 SNVs. Requires an internet connection, but try yourself.
# regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIB_BRU_K034", start = 2074, stop=3585)
# HIVmix <- deepSNV(test = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/test.bam", control = "http://
data(HIVmix) # attach data instead..
show(HIVmix)
plot(HIVmix)
head(summary(HIVmix))
```

RCC

*Example RCC data...***Description**

Example RCC data

**Details**

Deep sequencing experiments of a renal cell carcinoma and healthy control tissue.

**Examples**

```
data("RCC", package="deepSNV")
summary(RCC, adjust.method="bonferroni")[,1:6]
plot(RCC)
RCC.bb <- estimateDispersion(RCC, alternative="two.sided")
summary(RCC.bb, adjust.method="bonferroni")[,1:6]
plot(RCC.bb)
```

repeatMask-methods

*Mask homopolymeric repeats.***Description**

Mask homopolymeric repeats.

**Usage**

```
## S4 method for signature 'DNAStrng'
repeatMask(x, w=5, flank=TRUE)
## S4 method for signature 'deepSNV'
repeatMask(x, w=5, flank=TRUE)
```

**Arguments**

x	An object. Either a <a href="#">deepSNV-class</a> object or a <a href="#">DNAStrng</a> with the nucleotide sequence.
flank	Boolean. Indicates whether the sites adjacent to the repeat should also be masked.
w	Integer. The minimal length at which repeats should be masked. Default w=0.

**Details**

repeatMask-methods: This function masks homopolymeric repeats longer than a given width. These are hot-spots of sequencing error and can confound the analysis.

**Value**

repeatMask-methods: A boolean vector where TRUE indicates a non-homopolymeric region.

**Author(s)**

gemoritz

**Examples**

```
data(HIVmix)
which(repeatMask(HIVmix))
```

---

 RF

*Relative frequencies.*


---

**Description**

Relative frequencies.

**Usage**

```
RF(freq, total=FALSE)
```

**Arguments**

freq	A matrix with nucleotide counts.
total	If the nucleotide counts have columns for forward and reverse direction, return each strand separately (FALSE), or add the two (TRUE).

**Details**

Convenience function to compute the relative frequencies from a matrix with absolute counts.

**Value**

A matrix with the relative frequencies.

**Author(s)**

gemoritz

**Examples**

```
data(HIVmix)
RF(test(HIVmix))[1:10,]
RF(test(HIVmix), total=TRUE)[1:10,]
```

---

show,deepSNV-method    *Show method for deepSNV objects...*

---

## Description

Show method for deepSNV objects

## Usage

```
## S4 method for signature 'deepSNV'  
show(object)
```

## Arguments

object            A [deepSNV-class](#) object.

## Value

NULL

## Author(s)

gemoritz

## Examples

```
## Short example with 2 SNVs at frequency ~10%  
regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIB_BRU_K034", start = 3120, stop=3140)  
ex <- deepSNV(test = system.file("extdata", "test.bam", package="deepSNV"), control = system.file("extdata/  
show(ex)    # show method  
plot(ex)    # scatter plot  
summary(ex)    # summary with significant SNVs  
ex[1:3,]    # subsetting the first three genomic positions  
tail(test(ex, total=TRUE))    # retrieve the test counts on both strands  
tail(control(ex, total=TRUE))  
  
## Not run: Full example with ~ 100 SNVs. Requires an internet connection, but try yourself.  
# regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIB_BRU_K034", start = 2074, stop=3585)  
# HIVmix <- deepSNV(test = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/test.bam", control = "http://  
data(HIVmix) # attach data instead..  
show(HIVmix)  
plot(HIVmix)  
head(summary(HIVmix))
```

---

summary-methods      *Summary of a deepSNV object...*

---

## Description

Summary of a deepSNV object

## Usage

```
## S4 method for signature 'deepSNV'
summary(object, sig.level=0.05, adjust.method="bonferroni", fold.change=1)
```

## Arguments

object	A <a href="#">deepSNV-class</a> object.
sig.level	The desired significance level.
adjust.method	The adjustment method for multiple testing corrections. See <a href="#">p.adjust</a> for details. Set to NULL, for no adjustment. Default "bonferroni".
fold.change	The minimal fold change required of the relative frequency. Default 1.

## Details

summary-methods: Tabularize significant SNVs by evaluating the p-values of the [deepSNV](#) test.

## Value

summary-methods: A data.frame with the following columns:

chr	The chromosome
pos	The position (1-based)
ref	The reference (consensus) nucleotide
var	The variant nucleotide
p.val	The (corrected) p-value
freq.var	The relative frequency of the SNV
sigma2.freq.var	The estimated variance of the frequency
n.tst.fw	The variant counts in the test experiment, forward strand
cov.tst.fw	The coverage in the test experiment, forward strand
n.tst.bw	The variant counts in the test experiment, backward strand
cov.tst.bw	The coverage in the test experiment, backward strand
n.ctrl.fw	The variant counts in the control experiment, forward strand
cov.ctrl.fw	The coverage in the control experiment, forward strand
n.ctrl.bw	The variant counts in the control experiment, backward strand
cov.ctrl.bw	The coverage in the control experiment, backward strand
raw.p.val	The raw p-value

**Author(s)**

gemoritz

**Examples**

```
## Short example with 2 SNVs at frequency ~10%
regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIB_BRU_K034", start = 3120, stop=3140)
ex <- deepSNV(test = system.file("extdata", "test.bam", package="deepSNV"), control = system.file("extdata", "control.bam", package="deepSNV"))
show(ex) # show method
plot(ex) # scatter plot
summary(ex) # summary with significant SNVs
ex[1:3,] # subsetting the first three genomic positions
tail(test(ex, total=TRUE)) # retrieve the test counts on both strands
tail(control(ex, total=TRUE))

## Not run: Full example with ~ 100 SNVs. Requires an internet connection, but try yourself.
# regions <- data.frame(chr="B.FR.83.HXB2_LAI_IIIB_BRU_K034", start = 2074, stop=3585)
# HIVmix <- deepSNV(test = "http://www.bsse.ethz.ch/cbg/software/deepSNV/data/test.bam", control = "http://www.bsse.ethz.ch/cbg/software/deepSNV/control/control.bam")
data(HIVmix) # attach data instead..
show(HIVmix)
plot(HIVmix)
head(summary(HIVmix))
```

test-methods

*Get test counts...***Description**

Get test counts

**Usage**

```
## S4 method for signature 'deepSNV'
test(deepSNV, total=F)
```

**Arguments**

```
deepSNV      a deepSNV-class object
total        Logical. If true the sum of both strands is returned
```

**Details**

test-methods: Convenience function to obtain the test counts from a deepSNV object.

**Value**

test-methods: A matrix with the absolute frequencies summed over both strands.

**Examples**

```
data(HIVmix)
test(HIVmix)[1:10,]
test(HIVmix, total=TRUE)[1:10,]
```

---

`trueSNVs`*Example...*

---

**Description**

Example .bam data and true SNVs.

**Details**

Two .bam alignments as example data sets are downloaded remotely via http. Sequenced were a 1,512 nt fragment of the HIV genome and a mixture (90% + 10%) with another variants. The two sequences were confirmed by Sanger sequencing and stored in the table trueSNVs.

**Examples**

```
data(HIVmix)
data(trueSNVs)
table(p.adjust(p.val(HIVmix), method="BH") < 0.05, trueSNVs)
```

# Index

## \*Topic **package**

- deepSNV-package, 2
- [, 7
- [ (Extract-methods), 11
- [, deepSNV, ANY, ANY-method
  - (Extract-methods), 11
- [, deepSNV-method (Extract-methods), 11
  
- bam2R, 3, 9
  
- call, 7
- character, 7
- consensusSequence, 7
- consensusSequence
  - (consensusSequence-methods), 4
- consensusSequence, deepSNV-method
  - (consensusSequence-methods), 4
- consensusSequence, matrix-method
  - (consensusSequence-methods), 4
- consensusSequence-methods, 4
- control, 7
- control (control-methods), 5
- control, deepSNV-method
  - (control-methods), 5
- control-methods, 5
- coordinates, 7
- coordinates (coordinates-methods), 6
- coordinates, deepSNV-method
  - (coordinates-methods), 6
- coordinates-methods, 6
  
- data.frame, 6, 7, 9
- deepSNV, 3, 7, 9, 12, 20
- deepSNV (deepSNV-methods), 8
- deepSNV, character, character-method
  - (deepSNV-methods), 8
- deepSNV, character, matrix-method
  - (deepSNV-methods), 8
- deepSNV, deepSNV, missing-method
  - (deepSNV-methods), 8
- deepSNV, matrix, character-method
  - (deepSNV-methods), 8
- deepSNV, matrix, matrix-method
  - (deepSNV-methods), 8
  
- deepSNV-class, 6
- deepSNV-methods, 8
- deepSNV-package, 2
- DNAStrng, 5, 17
  
- estimateDirichlet, 7, 9
- estimateDirichlet
  - (estimateDirichlet-methods), 10
- estimateDirichlet, deepSNV-method
  - (estimateDirichlet-methods), 10
- estimateDirichlet, matrix-method
  - (estimateDirichlet-methods), 10
- estimateDirichlet-methods, 10
- estimateDispersion, 7, 9
- estimateDispersion
  - (estimateDispersion-methods), 10
- estimateDispersion, deepSNV, missing-method
  - (estimateDispersion-methods), 10
- estimateDispersion, matrix, matrix-method
  - (estimateDispersion-methods), 10
- estimateDispersion-methods, 10
- Extract-methods, 11
  
- GRanges, 9
  
- HIVmix (trueSNVs), 22
  
- list, 7
- loess, 13
  
- manhattanPlot, 12
- matrix, 4, 6, 7
  
- normalize, 7
- normalize (normalize-methods), 13
- normalize, deepSNV, missing-method
  - (normalize-methods), 13
- normalize, matrix, matrix-method
  - (normalize-methods), 13
- normalize-methods, 13
- numeric, 7

- p.adjust, [20](#)
- p.combine, [8](#), [14](#)
- p.val, [7](#)
- p.val (p.val-methods), [15](#)
- p.val, deepSNV-method (p.val-methods), [15](#)
- p.val-methods, [15](#)
- phiX, [15](#)
- plot.deepSNV, [16](#)
  
- RCC, [17](#)
- repeatMask, [7](#)
- repeatMask (repeatMask-methods), [17](#)
- repeatMask, deepSNV-method
  - (repeatMask-methods), [17](#)
- repeatMask, DNASTring-method
  - (repeatMask-methods), [17](#)
- repeatMask-methods, [17](#)
- RF, [18](#)
  
- show, [7](#)
- show (show, deepSNV-method), [19](#)
- show, deepSNV-method, [19](#)
- summary, [7](#)
- summary (summary-methods), [20](#)
- summary, deepSNV-method
  - (summary-methods), [20](#)
- summary-methods, [20](#)
  
- test, [7](#)
- test (test-methods), [21](#)
- test, deepSNV-method (test-methods), [21](#)
- test-methods, [21](#)
- trueSNVs, [22](#)
  
- vector, [7](#)