

Package ‘falconx’

February 24, 2017

Version 0.2

Date 2017-02-16

Title Finding Allele-Specific Copy Number in Whole-Exome Sequencing Data

Author Hao Chen and Nancy R. Zhang

Maintainer Hao Chen <hxchen@ucdavis.edu>

Depends R (>= 3.0.1)

Description This is a method for Allele-specific DNA Copy Number profiling for whole-Exome sequencing data. Given the allele-specific coverage and site biases at the variant loci, this program segments the genome into regions of homogeneous allele-specific copy number. It requires, as input, the read counts for each variant allele in a pair of case and control samples, as well as the site biases. For detection of somatic mutations, the case and control samples can be the tumor and normal sample from the same individual. The implemented method is based on the paper: Chen, H., Jiang, Y., Maxwell, K., Nathanson, K. and Zhang, N. (under review). Allele-specific copy number estimation by whole Exome sequencing.

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2017-02-24 23:43:44

R topics documented:

biasMatrix	2
falconx	2
getASCN.x	3
getChangepoints.x	4
pos	5
readMatrix	5
tauhat	6
view	6

Index	8
--------------	----------

 biasMatrix

Bias Matrix

Description

This is a data frame with two columns and the column names are "sN", "sT". They are the site-specific bias in total coverage for normal (control) sample and tumor (case) sample, respectively.

 falconx

Finding Allele-Specific Copy Number in Whole-Exome Sequencing Data

Description

This library contains a set of tools for allele-specific DNA copy number profiling using whole exome sequencing. Given the allele-specific coverage and site biases at the variant loci, this program segments the genome into regions of homogeneous allele-specific copy number. It requires, as input, the read counts for each variant allele in a pair of case and control samples, as well as the site biases. For detection of somatic mutations, the case and control samples can be the tumor and normal sample from the same individual.

Author(s)

Hao Chen and Nancy R. Zhang

Maintainer: Hao Chen (hxchen@ucdavis.edu)

See Also

[getChangepoints.x](#), [getASCN.x](#), [view](#)

Examples

```
data(Example)
# tauhat = getChangepoints.x(readMatrix, biasMatrix) # uncomment this to run the function.
cn = getASCN.x(readMatrix, biasMatrix, tauhat=tauhat)
# cn$tauhat would give the indices of change-points.
# cn$ascn would give the estimated allele-specific copy numbers for each segment.
# cn$Haplotype[[i]] would give the estimated haplotype for the major chromosome in segment i
# if this segment has different copy numbers on the two homologous chromosomes.
view(cn)
```

Description

Given a set of breakpoints where parent-specific copy number changes, this function estimates the parent-specific copy number for each segment, and the haplotype for the major chromosome on segments where the two homologous chromosomes have different copy numbers. You are recommended to specify the parameter "rdep", the case-control genome-wide average coverage ratio. Usually, a good estimate of rdep is (total mapped reads in tumor)/(total mapped reads in normal).

Usage

```
getASCN.x(readMatrix, biasMatrix, tauhat=NULL, threshold=0.15, COri=c(0.95,1.05),
error=1e-5, maxIter=1000, independence=TRUE, pos=NULL, readlength=NULL)
```

Arguments

readMatrix	A data frame with four columns and the column names are "AN", "BN", "AT" and "BT". They are A-allele coverage in the tumor (case) sample, B-allele coverage in the tumor (case) sample, A-allele coverage in the normal (control) sample, and B-allele coverage in the normal (control) sample, respectively.
biasMatrix	A data frame with two columns and the column names are "sN", "sT". They are the site-specific bias in total coverage for normal (control) sample and tumor (case) sample, respectively.
tauhat	The estimated break points. If it is not specified (NULL), then this function will first estimate the break points by calling the function "getChangepoints.x", and then estimate the parent-specific DNA copy number for each segment.
threshold	The estimated copy number are set to be 1 if it differs from 1 by less than this threshold.
COri, error, maxIter	Parameters used in estimating the success probabilities of the mixed binomial distribution. See the manuscript by Chen and Zhang for more details. "pOri" provides the initial success probabilities. The two values in pOri needs to be different. "error" provides the stopping criterion. "maxIter" is the maximum iterating steps if the stopping criterion is not achieved.
independence	The model assumes reads are conditionally independent. If this argument is FALSE, the pruning approach will be performed.
pos	The locations (in base pair) of the heterozygous sites. This information is needed when "independence=FALSE".
readlength	The length of read if the data is from single-end sequencing, and the maximum span of read pairs if the data if from paired-end sequencing. This information is needed when "independence=FALSE".

Value

tauhat	A vector holding the estimated break points in terms of the index in the coverage vectors.
ascn	The estimated parent-specific copy numbers in the segments between the break points in tauhat.
Haplotype	The estimated haplotype for the major chromosome (the chromosome has a higher copy number compared to its homologous chromosome) on segments where the two homologous chromosomes have different copy numbers.

See Also

[getChangepoints.x, view](#)

Examples

```
data(Example)
cn = getASCN.x(readMatrix, biasMatrix, tauhat=tauhat)
# cn$tauhat would give the indices of change-points.
# cn$ascn would give the estimated allele-specific copy numbers for each segment.
# cn$Haplotype[[i]] would give the estimated haplotype for the major chromosome in segment i
# if this segment has different copy numbers on the two homologous chromosomes.
```

getChangepoints.x *Getting Change-points*

Description

This function estimates the change-points where one or both parent-specific copy numbers change. It uses a circular binary segmentation approach to find change-points in a binomial mixture process. The output of the function is the set of locations of the break points. If the whole genome is analyzed, it is recommended to run this function chromosome by chromosome, and the runs on different chromosomes can be done in parallel to shorten the running time.

Usage

```
getChangepoints.x(readMatrix, biasMatrix, verbose=TRUE, COri=c(0.95,1.05), error=1e-5,
maxIter=1000, independence=TRUE, pos=NULL, readlength=NULL)
```

Arguments

readMatrix	A data frame with four columns and the column names are "AN", "BN", "AT" and "BT". They are A-allele coverage in the tumor (case) sample, B-allele coverage in the tumor (case) sample, A-allele coverage in the normal (control) sample, and B-allele coverage in the normal (control) sample, respectively.
biasMatrix	A data frame with two columns and the column names are "sN", "sT". They are the site-specific bias in total coverage for normal (control) sample and tumor (case) sample, respectively.

verbose	Provide progress messages if it is TRUE. This argument is TRUE by default. Set it to be FALSE if you want to turn off the progress messages.
COri, error, maxIter	Parameters used in estimating the success probabilities of the mixed binomial distribution. See the manuscript by Chen and Zhang for more details. "pOri" provides the initial success probabilities. The two values in pOri needs to be different. "error" provides the stopping criterion. "maxIter" is the maximum iterating steps if the stopping criterion is not achieved.
independence	The model assumes reads are conditionally independent. If this argument is FALSE, the pruning approach will be performed.
pos	The locations (in base pair) of the heterozygous sites. This information is needed when "independence=FALSE".
readlength	The length of read if the data is from single-end sequencing, and the maximum span of read pairs if the data if from paired-end sequencing. This information is needed when "independence=FALSE".

See Also

[getASCN.x](#)

Examples

```
data(Example)
# tauhat = getChangepoints.x(readMatrix, biasMatrix) # uncomment this to run the function.
```

pos	<i>Position (bp)</i>
-----	----------------------

Description

This is a vector containing position (in base pair) of each heterozygous site for the read count data in "Example.rda".

readMatrix	<i>Reads Matrix</i>
------------	---------------------

Description

This is a data frame with four columns and the column names are "AN", "BN", "AT" and "BT". They are A-allele coverage in the tumor (case) sample, B-allele coverage in the tumor (case) sample, A-allele coverage in the normal (control) sample, and B-allele coverage in the normal (control) sample, respectively.

tauhat	<i>Estimated Break Points</i>
--------	-------------------------------

Description

This is a vector containing the estimated break points that one would get by calling the function `getChangepoints.x(readMatrix, biasMatrix)` with "readMatrix" and "biasMatrix" from "Example.rda".

view	<i>Viewing Data with Allele-specific Copy Number</i>
------	------------------------------------------------------

Description

This function generates three plots: The first plots the A-allele frequencies of the case (black) sample overlaid onto those of the control (gray) sample; the second plots the relative depth of the case over control adjusted by the ratio of total mapped reads, i.e. $P \cdot (\text{read count in tumor}) / (\text{read count in normal})$, where $P = (\text{total reads mapped in normal}) / (\text{total reads mapped in tumor})$; the third plots the estimated parent-specific DNA copy numbers.

Usage

```
view(output, pos=NULL, rdep=NULL, plot="all", independence=TRUE, ...)
```

Arguments

output	The output from calling function "getASCN.x".
pos	A vector of the base positions for the SNPs. If this information is not provided, the x-axis of the plots will simply be the SNP ordering. If this information is provided, the x-axis of the plots will be the position information.
rdep	The relative depth of the case sample over the control sample. If it is not specified (NULL), then the value $\text{median}(AT+BT) / \text{median}(AN+BN)$ will be used.
plot	This argument determines what to plot. By default, this function gives all three plots described above ("all"). You can also plot each one individually if you set this argument to either of "Afreq", "RelativeCoverage" or "ASCN".
independence	If argument "pos" is specified, when "independence=FALSE", the pruned positions will be used.
...	Arguments from plot can be passed along.

See Also

[getASCN.x](#)

Examples

```
data(Example)
cn = getASCN.x(readMatrix, biasMatrix, tauhat=tauhat)
view(cn)

# to view with position as the x-axis
view(cn, pos=pos)

# to view the plot for only showing A-allele frequency of the case (black) sample overlaid
# onto those of the control (gray) sample
par(mfrow=c(1,1))
view(cn, plot="Afreq")

# to view the relative depth of the case over control adjusted by the ratio of total mapped
# reads in fixed size bins
par(mfrow=c(1,1))
view(cn, plot="RelativeCoverage")

# to view the estimated allele-specific DNA copy numbers
par(mfrow=c(1,1))
view(cn, plot="ASCN")
```

Index

`biasMatrix`, [2](#)

`falconx`, [2](#)

`getASCN.x`, [2](#), [3](#), [5](#), [6](#)

`getChangepoints.x`, [2](#), [4](#), [4](#)

`pos`, [5](#)

`readMatrix`, [5](#)

`tauhat`, [6](#)

`view`, [2](#), [4](#), [6](#)