

Package ‘falcon’

October 13, 2022

Version 0.2

Date 2016-04-20

Title Finding Allele-Specific Copy Number in Next-Generation 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 using Next-Generation Sequencing. Given the allele-specific coverage 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. For detection of somatic mutations, the case and control samples can be the tumor and normal sample from the same individual.

License GPL (>= 2)

NeedsCompilation yes

Repository CRAN

Date/Publication 2016-04-21 08:20:42

R topics documented:

| | |
|---------------------------|---|
| falcon | 2 |
| getASCN | 3 |
| getChangepoints | 4 |
| pos | 5 |
| readMatrix | 5 |
| tauhat | 5 |
| view | 6 |

| | |
|--------------|----------|
| Index | 8 |
|--------------|----------|

falcon

Finding Allele-Specific Copy Number in Next-Generation Sequencing Data

Description

This library contains a set of tools for Allele-specific DNA Copy Number Profiling using Next-Generation Sequencing. Given the allele-specific coverage at a set of variant loci, `getChangepoints` program segments the genome into regions of homogeneous allele-specific copy number. It requires, as input, the read counts supporting each variant allele in each of a pair of case and control samples. For example, for detection of somatic mutations in a tumors, the case and control samples can be, respectively, the tumor and normal sample from the same individual.

Author(s)

Hao Chen and Nancy Zhang

Maintainer: Hao Chen (hxchen@ucdavis.edu)

References

Hao Chen, John Bell, Nicolas Zavala, Hanlee Ji and Nancy Zhang. Allele-specific copy number profiling by next-generation DNA sequencing. *Nucleic Acids Research*, 2014.

See Also

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

Examples

```
data(Example)
# tauhat = getChangepoints(readMatrix) # uncomment this to run the function.
# This example has 6309 variant loci and it took 94 seconds to run on a laptop with
# Intel Core i5-2410M processor.
cn = getASCN(readMatrix, 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)
```

getASCN

*Getting Allele-specific DNA Copy Number***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(readMatrix, rdep=NULL, tauhat=NULL, threshold=0.15, pOri=c(0.49,0.51),
error=1e-5, maxIter=1000)
```

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. |
| rdep | The case-control coverage ratio, i.e., the ratio of the total number of mapped reads in case sample versus that in the control sample. If this is not specified (NULL), then the value $\text{median(AT+BT)}/\text{median(AN+BN)}$ will be used. |
| tauhat | The estimated break points. If it is not specified (NULL), then this function will first estimate the break points by calling the function "gettau", 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. |
| pOri, 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. |

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](#), [view](#)

Examples

```
data(Example)
cn = getASCN(readMatrix, 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 *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(readMatrix, verbose=TRUE, pOri=c(0.49,0.51), error=1e-5, maxIter=1000)
```

Arguments

| | |
|----------------------|--|
| readMatrix | A data frame with four columns: AN, BN, AT, BT. |
| 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. |
| pOri, 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. |

References

Hao Chen, John Bell, Nicolas Zavala, Hanlee Ji and Nancy Zhang. Allele-specific copy number profiling by next-generation DNA sequencing. *Nucleic Acids Research*, 2014.

See Also

[getASCN](#)

Examples

```
data(Example)
# tauhat = getChangepoints(readMatrix) # uncomment this to run the function.
# This example has 6309 variant loci and it took 94 seconds to run on a laptop with
# Intel Core i5-2410M processor.
```

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

Description

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

| | |
|------------|------------------------------------|
| readMatrix | <i>An example reads count data</i> |
|------------|------------------------------------|

Description

This is a toy data set we created by taking a subset of the heterozygous SNP sites from data generated by Illumina deep sequencing of a pair of tumor and matched normal samples. The table has four columns with the column names "AN", "BN", "AT" and "BT", which are the 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. Coverage is defined as simply the number of reads containing the allele among all reads that cover the site. Each row corresponds to one variant site. Their corresponding positions in the reference genome are listed in vector "pos" in "Example.rda". The order of the variant sites is the same order as they appear on the chromosome of the reference genome.

| | |
|--------|-------------------------------|
| 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(readMatrix)` with "readMatrix" from the "Example.rda".

 view

Viewing Data with Allele-specific Copy Number

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", ...)
```

Arguments

| | |
|--------|---|
| output | The whole output from calling function "getASCN". |
| 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". |
| ... | Arguments from plot can be passed along. |

See Also

[getASCN](#)

Examples

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

# 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")
```

view

7

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

Index

falcon, [2](#)

getASCN, [2](#), [3](#), [4](#), [6](#)

getChangepoints, [2](#), [4](#), [4](#)

pos, [5](#)

readMatrix, [5](#)

tauhat, [5](#)

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