Package 'SomatiCA'

April 10, 2015

Type Package

Title SomatiCA: identifying, characterizing, and quantifying somatic copy number aberrations from cancer genome sequencing

Version 1.10.0

Date 2013-10-29

Author Mengjie Chen <mengjie.chen@yale.edu>, Hongyu Zhao <hongyu.zhao@yale.edu>

Maintainer Mengjie Chen <mengjie.chen@yale.edu>

Imports foreach, lars, sn, DNAcopy, methods, rebmix, GenomicRanges, IRanges

Depends R (>= 2.14.0), lars, DNAcopy, foreach, methods, rebmix, GenomicRanges, IRanges, doParallel

Enhances sn, SomatiCAData

Description SomatiCA is a software suite that is capable of identifying, characterizing, and quantifying somatic CNAs from cancer genome sequencing. First, it uses read depths and lesser allele frequencies (LAF) from mapped short sequence reads to segment the genome and identify candidate CNAs. Second, SomatiCA estimates the admixture rate from the relative copy-number profile of tumor-normal pair by a Bayesian finite mixture model. Third, SomatiCA quantifies absolute somatic copy-number and subclonality for each genomic segment to guide its characterization. Results from SomatiCA can be further integrated with single nucleotide variations (SNVs) to get a better understanding of the tumor evolution.

License GPL (>=2)

biocViews Sequencing, CopyNumberVariation

R topics documented:

`SomatiCA-package` *Identifying, characterizing and quantifying somatic copy number aberrations from cancer genome sequencing*

Description

SomaticCNA is a software suite that is capable of identifying, characterizing, and quantifying somatic CNAs from cancer genome sequencing. First, it uses read depths and lesser allele frequencies (LAF) from mapped short sequence reads to segment the genome and identify candidate CNAs. Second, SomaticCNA estimates the admixture rate from the relative copy-number profile of tumor-normal pair by a Bayesian finite mixture model. Third, SomaticCNA quantifies absolute somatic copy-number and subclonality for each genomic segment to guide its characterization.

Details

Package:	SomatiCA
Type:	Package
Version:	0.99.1
Date:	2012-11-13
License:	GPL (>=2)

Author(s)

Mengjie Chen, Hongyu Zhao Maintainer: Mengjie Chen <mengjie.chen@yale.edu>

See Also

segment

Examples

```
rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05), rnorm(200, 0.2, 0.05), rnorm(200,
rawLAF <- ifelse(rawLAF>0.5, 1-rawLAF, rawLAF)
germLAF <- c(rnorm(800+650, 0.4, 0.05))
germLAF <- ifelse(germLAF>0.5, 1-germLAF, germLAF)
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60), rpois(200, 25), rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650,50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(c(1:800), c(1:650))
zygo <- rep("het", 800+650)
x <- data.frame(chr, as.integer(position), as.character(zygo), as.integer(reads1), rawLAF, as.integer(reads2), ge
colnames(x) <- c("seqnames", "start", "zygosity", "tCount", "LAF", "tCountN", "germLAF")
data <- SomatiCAFormat(x)

### This is an easy example, without much noise.
### Consider to use rss=T to select change points from sequencing data
seg <- larsCBSsegment(data, rss = FALSE)

plotSegment(seg$segment, data, k = 2, smooth = FALSE)
```

---

admixtureRate *Estimate the admixture rate of normal cells in a tumor sample.*

Description

The estimation of the admixture rate is accomplished by fitting the input tumor somatic copy number (somatic ratio*2) of all segments with a Bayesian finite mixture model, with components centered at the discrete levels. Each segment was assigned with a discrete level based on corresponding posterior probability. Segments with ambiguous assignments will be classified as candidate subclonal events and excluded from admixture rate inference. The admixture rate will be estimated by an optimal solution contributed by explanation of tumor copy number with all remaining segments as integer level.

Usage

```
admixtureRate(segmentwithratio, mcmc = 10000, burnin = 5000, p = 0.01, weight=FALSE)
```

Arguments

segmentwithratio	A GRanges object, segments with annotation of somatic ratio.
mcmc	number of MCMC iteration.
burnin	number of MCMC iteration for burnin.
p	posterior probability cutoff for ambiguous integer copy number assignments.
weight	whether penalize the segment based on its length.

Value

admix	Admixture rate.
mu	Posterior mean for each discrete level.
cluster	A two-column matrix for input somatic ratio and corresponding integer somatic copy number level.

Author(s)

Mengjie Chen

Examples

```
data(segwithratio)
seg <-  GRanges(seqnames=segwithratio$chromosome,
              ranges=IRanges(start=segwithratio$start,
                             end=segwithratio$end),
              medLAF=segwithratio$medLAF,
              medgLAF=segwithratio$medgermlineLAF,
              ratio=segwithratio$ratio)
bb <- admixtureRate(seg)
```

---

collapse *Collapse measurements with a certain bin size.*

Description

Collapse measurements with a certain bin size.

Usage

```
collapse(data, k = 5)
```

Arguments

data	A numeric vector.
k	Bin size.

Value

A numeric vector.

Author(s)

Mengjie Chen

Examples

```
x <- rnorm(500)
y <- collapse(x, k = 5)
```

`copynumberCorrected` *Somatic copy number corrected by admixture rate.*

Description

Somatic copy number corrected by admixture rate.

Usage

```
copynumberCorrected(segment, admix)
```

Arguments

`segment`	A GRanges class, segments with annotation of somatic ratio.
`admix`	Numeric. Admixture rate of normal cells.

Value

A GRanges class, segments with annotation of event.

`seqnames, ranges`
: See GRanges().

`medLAF`
: A numeric vector. Median lesser allele frequency of each segment in the tumor sample.

`ratio`
: A numeric vector. Somatic ratio (read depth ratio of tumor/normal) of each segment.

`somaCN`
: An integer vector. Somatic integer copy number based on somatic ratio.

`event`
: A character vector. Somatic event characterization based on somatic ratio, taking values of "=", "Loss", "Gain", "LOH", "neutral LOH" or "double deletion".

Author(s)

Mengjie Chen

Examples

```
data(segwithratio)
x <-  GRanges(seqnames=segwithratio$chromosome,
              ranges=IRanges(start=segwithratio$start,
                             end=segwithratio$end),
              medLAF=segwithratio$medLAF,
              medgLAF=segwithratio$medgermlineLAF,
              ratio=segwithratio$ratio
              )
admix <- 0.4
y <- copynumberCorrected(x, admix)
```

`denoise` *Smoothing procedure by replacing outliers.*

Description

Smoothing procedure by replacing outliers (defined by deviations from mean) with median in a sliding window.

Usage

```
denoise(data, k = 30, t = 2)
```

Arguments

`data`	A numeric vector.
`k`	Define window size.
`t`	The number of standard deviation used to define outliers.

Value

Smoothed data, a numeric vector.

Author(s)

Mengjie Chen

Examples

```
x <- c(rnorm(200, 0.4, 0.05), rnorm(200, 0.1, 0.05))
x[round(runif(20)*400)] <- rnorm(20, 0.2, 0.05)
x[round(runif(20)*400)] <- rnorm(20, 0.3, 0.05)
y <- denoise(x)
plot(x)
points(y, pch=20, col="red")
```

GCbiasRemoval *Correct GC bias for read depth ratio of each site.*

Description

Read depth ratio is defined as read depth at each site divided by median of read depth of that sequencing library. SomatiCA corrects GC bias for read depth ratio at each site based on a linear model described in Diskin et al.(2008).

Usage

```
GCbiasRemoval(input, GCcontent)
```

Arguments

input	A GRanges object. Same as the output of SomatiCAFormat().
GCcontent	A data frame object with 4 column."chr", "interval1", "interval2" and "GC".

Value

A GRanges object.

zygosity, tCount, LAF, tCountN, germLAF
: Same as the outout SomatiCAFormat()

tRcorrected
: GC bias corrected read depth ratio for tumor sample.

nRcorrected
: GC bias corrected read depth ratio for normal sample.

Author(s)

Mengjie Chen

References

Diskin et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Research*, 36(19):e126, 2008.

See Also

See Also segmentGCbiasRemoval.

Examples

```
rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05), rnorm(200, 0.2, 0.05), rnorm(200,
germLAF <- c(rnorm(800+650, 0.4, 0.05))
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60),  rpois(200, 25), rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650, 50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(seq(1, 1600000, by=2000), seq(1, 1300000, by=2000))
```

```
zygo <- rep("het", 800+650)
data <- GRanges(seqnames=chr,
        ranges=IRanges(start=position, width=1),
        zygosity=zygo,
        tCount=reads1,
        LAF=rawLAF,
        tCountN=reads2,
        germLAF=germLAF)
data(GCcontent)
x <- GCbiasRemoval(data, GCcontent)
```

GCcontent *GC content for hg19 human genome at each 1Mb interval.*

Description

GC content was calculated for hg19 human genome at each 1Mb interval. SomatiCA uses this file to correct GC bias.

Usage

```
data(GCcontent)
```

Format

A data frame with 2897 rows on the following 4 variables.

`chr` A character vector.

`interval1` A numeric vector.

`interval2` A numeric vector.

`GC` A numeric vector. GC content.

Source

Assembly human genome (hg19, GRCh37 Genome Reference Consortium Human Reference 37 (GCA_000001405.1)) was downloaded from http://hgdownload.soe.ucsc.edu/goldenPath/hg19/chromosomes/.

GCcount *Calculate the GC content of given window size for a chromosome.*

Description

The function downloads the .fa.gz of a given chromosome from UCSC genome browser and calculate the GC content for given window size.

Usage

```
GCcount(chr, binsize, url)
```

Arguments

chr	chromosome name. e.g. "chr1".
binsize	window size. e.g. 1000000.
url	diretory of human genome assembly. e.g."http://hgdownload.soe.ucsc.edu/goldenPath/hg19/chromosome

Value

A GRanges object.

chr	A character vector. Chromosome information.
interval1	An integer vector. Start position of each bin.
interval2	An integer vector. End position of each bin.
GC	A numeric vector. GC content of each bin.

Author(s)

Mengjie Chen

Examples

```
chr <- "chr1"
url <- "http://hgdownload.soe.ucsc.edu/goldenPath/hg19/chromosomes/"
#downloding speed may depend on the machine and internet
#test <- GCcount(chr, 10000, url)
```

`larsCBSsegment` *Segmentation based on Circular Binary Segmentation followed by a model selection procedure on detected change points.*

Description

A model selection procedure is applied after CBS segmentation. In another word, we assess which ones in over-detected change points from CBS calls are really necessary. More specifically, we used K change points as K predictors for input $X_i, i = (0,..., n)$ to fit a linear model and select variables by step-wise regression implemented in lars()(from R package lars). Then optimal change points could be selected from the LARS solution path via different criterions.

Usage

```
larsCBSsegment(data, selection = .selection.default(), collapse.k = 0, ncores = 1, verbose = TRUE, varia
```

Arguments

`data`	A GRanges object, output of SomatiCAFormat().
`selection`	Model selection parameters.
`collapse.k`	Number of data points collapsed.
`ncores`	Number of cores used.
`verbose`	Whether working messages are shown.
`variation.control`	A logical value, whether pseudo points are used to smooth the segment. Default is TRUE.
`rss`	A logical value, whether a cutoff based on residue sum of squares is used. Default is FALSE.
`S`	The cutoff based on residue sum of squares. Default is 0.1.
`k`	The window size used to smooth the outliers.
`...`	Arguments for segment() in DNAcopy package.

Value

`segment`	S4 class, "Segmented".
`hetsites`	Heterozygous sites used in segmentation, unsmoothed.

Author(s)

Mengjie Chen

References

Efron, Hastie, Johnstone and Tibshirani (2003) "Least Angle Regression" (with discussion) *Annals of Statistics*. Olshen, A. B., Venkatraman, E. S., Lucito, R., Wigler, M. (2004). Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5: 557-572. Venkatraman, E. S., Olshen, A. B. (2007) A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics* 23: 657-63.

See Also

See Also SomatiCAFormat, lars, segment.

Examples

```
rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05), rnorm(200, 0.2, 0.05), rnorm(200,
rawLAF <- ifelse(rawLAF>0.5, 1-rawLAF, rawLAF)
germLAF <- c(rnorm(800+650, 0.4, 0.05))
germLAF <- ifelse(germLAF>0.5, 1-germLAF, germLAF)
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60), rpois(200, 25), rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650,50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(c(1:800), c(1:650))
zygo <- rep("het", 800+650)
x <- data.frame(chr, as.integer(position), as.character(zygo), as.integer(reads1), rawLAF, as.integer(reads2), ge
colnames(x) <- c("seqnames", "start", "zygosity", "tCount", "LAF", "tCountN", "germLAF")
data <- SomatiCAFormat(x)

### This is an easy example, without much noise.
### Consider to use rss=T to select change points from sequencing data
seg <- larsCBSsegment(data, rss = FALSE)

plotSegment(seg$segment, data, k = 1, smooth = FALSE)
plotSegment(seg$segment, data, k = 2, smooth = FALSE)
```

`MergeSegment` *Merge neighboring segments with same somatic copy number and events.*

Description

Take segments with subclonality characterization as input and merge same events.

Usage

```
MergeSegment(segment)
```

Arguments

segment	A GRanges object, output of subclonality().

Value

A GRanges object, merged segments.

Author(s)

Mengjie Chen

See Also

See Also as subclonality, ~~~

Examples

```
### This is just a toy example.
chr <- c("chr1", "chr1", "chr1", "chr1", "chr2", "chr2", "chr2", "chr2")
start <- c(1, 41, 61 , 71, 1, 51, 71, 91)
end <- c(41, 61 , 71, 91, 51, 71, 91, 101)
medLAF <- c(0.15, 0.27, 0.4, 0.42, 0.4, 0.41, 0.42, 0.39)
medgermlineLAF <- rep(0.42, 8)
ratio <- c(0.5, 0.7, 1.1, 1, 1.1, 1, 1, 0.9)
copynumber <- c(1, 1, 2, 2, 2, 2, 2, 2)
event <- c("LOH", "Loss", "=", "=", "=", "=", "=", "=")
clonality <- c("Clonal", "subclonal_loss", "subclonal_gain", "=", "=", "=", "=", "subclonal_loss")
germlinecopynumber <- c(2, 2, 2, 2, 2, 2, 2, 2)
subclonecopynumber <- c(1, 1, 3, 2, 2, 2, 2, 1)
subpercent <- c(1, 0.4, 0.1, 0, 0, 0, 0, 0.1)
x <- GRanges(seqnames=chr,
             ranges=IRanges(start=start, end=end),
             medLAF=medLAF,
           medgLAF=medgermlineLAF,
           ratio=ratio,
             somaCN=copynumber,
             event=event,
           clonality=clonality,
           germCN=germlinecopynumber,
           subclonalCN=subclonecopynumber,
             subpercent=subpercent)
merged <- MergeSegment(x)
```

---

plotSegment *Plot segmentation.*

Description

Plot segmentation.

Usage

```
plotSegment(segment, input, k = 1, col1 = "orange", col2 = "blue", smooth = FALSE, dev.new = TRUE, ...)
```

Arguments

segment	A GRanges object, segmentation result.
input	A GRanges object, input for SomatiCA.
k	When there are multiple chromosomes, only kth will be plot each time.
col1	Color for data points.
col2	Color for segments.
smooth	Whether smoothed data is plotted. Default is T.
dev.new	Whether a new device will be opened.
...	Parameters in plot().

Value

A plot.

Author(s)

Mengjie Chen

Examples

```
#data#
rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05), rnorm(200, 0.2, 0.05), rnorm(200,
reads <- rep(50, 800+650)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(c(1:800), c(1:650))
zygo <- rep("het", 800+650)
input <- GRanges(seqnames=chr,
                 ranges=IRanges(start=position, width=1),
                 zygosity=zygo,
                 tCount=reads,
                 LAF=rawLAF,
                 tCountN=reads,
                 gLAF=rawLAF)

#segment#
chr <- c("chr1", "chr1", "chr1", "chr2", "chr2", "chr2")
start <- c(1, 300, 600, 1, 200, 400)
end <- c(300, 600, 800, 200, 400, 650)
medLAF <- c(0.2, 0.4, 0.3, 0.2, 0.3, 0.4)
medgermlineLAF <- rep(0.4, 6)
segment <- GRanges(seqnames=chr,
                   ranges=IRanges(start=start, end=end),
                   medLAF=medLAF,
                   medgLAF=medgermlineLAF)

## First chromosome
plotSegment(segment, input, k=1)
## Second chromosome
plotSegment(segment, input, k=2)
```

plotSubclonality *Plot clonality and somatic copy number.*

Description

Plot clonality and somatic copy number for a tumor sample.

Usage

```
plotSubclonality(segment, dev.new = TRUE, ...)
```

Arguments

segment	A GRanges object, segments with annotation of copy number and subclonality.
dev.new	Whether a new device will be opened.
...	Parameters in plot().

Value

A plot.

Author(s)

Mengjie Chen

Examples

```
chr <- c("chr1", "chr1", "chr1", "chr1", "chr2", "chr2", "chr2", "chr2")
start <- c(1, 41, 61 , 71, 1, 51, 71, 91)
end <- c(41, 61 , 71, 91, 51, 71, 91, 101)
medLAF <- c(0.15, 0.27, 0.4, 0.42, 0.4, 0.41, 0.42, 0.39)
medgermlineLAF <- rep(0.43, 8)
ratio <- c(0.5, 0.7, 1.1, 1, 1.1, 1, 1, 0.9)
copynumber <- c(1, 1, 2, 2, 2, 2, 2, 2)
event <- c("LOH", "Loss", "=", "=", "=", "=", "=", "=")
clonality <- c("Clonal", "subclonal_loss", "subclonal_gain", "=",
            "=", "=", "=", "subclonal_loss")
germlinecopynumber <- c(2, 2, 2, 2, 2, 2, 2, 2)
subclonecopynumber <- c(1, 1, 3, 2, 2, 2, 2, 1)
subpercent <- c(1, 0.4, 0.1, 0, 0, 0, 0, 0.1)
x <- GRanges(seqnames=chr,
             ranges=IRanges(start=start, end=end),
             medLAF=medLAF,
             medgLAF=medgermlineLAF,
           ratio=ratio,
             somaCN=copynumber,
             event=event,
           clonality=clonality,
           germCN=germlinecopynumber,
```

```
            subclonalCN=subclonecopynumber,
              subpercent=subpercent)
merged <- MergeSegment(x)
plotSubclonality(merged)
```

`refineSegment` *Refine the segmentation based on estimated somatic ratio.*

Description

Neighbor segments with difference in somatic ratio less than certain threshold will be merged together.

Usage

```
refineSegment(segmentwithratio, data, threshold1 = 0.01, threshold2 = 0.05, adjust = FALSE, method = "ml
```

Arguments

`segmentwithratio`	A GRanges object, segments with annotation of somatic ratio, usually the output of somaticRatio().
`data`	A GRanges object, input data from SomatiCAFormat().
`threshold1`	The threshold used to merge the segments based on median LAF. Default is 0.01.
`threshold2`	The threshold used to merge the segments based on somatic ratio. Default is 0.05.
`method`	Method used to estimate somatic ratio of given segments. For the "mle" method somatic ratio is estimated by a maximum likelihood approach. For the "mean" method, somatic ratio is estimated by the ratio between mean of tumor sample and normal sample. For the "geometric", somatic ratio is estimated by geometric mean of somatic ratios of all sites in a given segment.
`adjust`	Adjust the normal and tumor library and make their median equal.

Value

A GRanges object, refined segments with annotation of somatic ratio.

Author(s)

Mengjie Chen

See Also

See Also `somaticRatio`.

Examples

```
chr <- c("chr1", "chr1", "chr1", "chr2", "chr2", "chr2")
start <- c(1, 300, 600, 1, 200, 400)
end <- c(300, 600, 800, 200, 400, 650)
medLAF <- c(0.2, 0.4, 0.3, 0.2, 0.3, 0.4)
gLAF <- rep(0.4, 6)
seg <- GRanges(seqnames=chr,
                   ranges=IRanges(start=start, end=end),
                   medLAF=medLAF,
                   medgLAF=gLAF)


rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05),
            rnorm(200, 0.2, 0.05), rnorm(200, 0.3, 0.05), rnorm(250, 0.4, 0.05))
germLAF <- c(rnorm(800+650, 0.4, 0.05))
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60),  rpois(200, 25),
            rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650,50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(c(1:800), c(1:650))
zygo <- rep("het", 800+650)
data <- GRanges(seqnames=chr,
                 ranges=IRanges(start=position, width=1),
                 zygosity=zygo,
                 tCount=reads1,
                 LAF=rawLAF,
                 tCountN=reads2,
                 germLAF=germLAF)

x <- somaticRatio(seg, data, method = "mle")
y <- refineSegment(x, data)
```

`segmentGCbiasRemoval` *Correct GC bias for read depth ratio of each segment.*

Description

Read depth ratio is defined as read depth at each site divided by median of read depth of that sequencing library. SomatiCA corrects GC bias for read depth ratio at each site based on a linear model described in Diskin et al.(2008). The read depth ratio for each segment is calculated as the geometric mean of all sites in that segment.

Usage

```
segmentGCbiasRemoval(segment, input, GC, remove=TRUE)
```

Arguments

segment	A GRanges class, segments with annotation of integer copy number and event.
input	A GRanges class, SomatiCA input from SomatiCAFormat().
GC	A data frame object with 4 column."chr", "interval1", "interval2" and "GC".
remove	Whether GC bias needed to be removed at this step. Default is TRUE. If GC bias has been adjusted in pre-processing, then use FALSE.

Value

A GRanges class, segments with annotation of GC corrected read depth ratio.

chromosome, start, end, medLAF, ratio, copynumber, event	Same as the output of copynumberCorrected().
tRcorrected	GC bias corrected read depth ratio for tumor sample.
nRcorrected	GC bias corrected read depth ratio for normal sample.

Author(s)

Mengjie Chen

References

Diskin et al. Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms. *Nucleic Acids Research*, 36(19):e126, 2008.

See Also

See Also GCbiasRemoval.

Examples

```
### generate sequencing input ###
rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05), rnorm(200, 0.2, 0.05), rnorm(200,
germLAF <- c(rnorm(800+650, 0.4, 0.05))
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60),  rpois(200, 25), rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650, 50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(seq(1, 16000000, by=20000), seq(1, 13000000, by=20000))
zygo <- rep("het", 800+650)
data <- GRanges(seqnames=chr,
        ranges=IRanges(start=position, width=1),
        zygosity=zygo,
        tCount=reads1,
        LAF=rawLAF,
        tCountN=reads2,
        germLAF=germLAF)

### generate pseudo segments ###
```

```
chr <- c("chr1", "chr1", "chr1", "chr2", "chr2", "chr2")
start <- position[c(1, 301, 601, 1, 201, 401)]
end <-  position[c(301, 601, 800, 201, 401, 651)]
medLAF <- c(0.2, 0.4, 0.3, 0.2, 0.3, 0.4)
gLAF <- rep(0.43, 6)
ratio <- c(0.5, 1, 1.3, 0.5, 0.8, 1)
copynumber <- c(1, 2, 3, 1, 3, 2)
event <- c("LOH", "=", "Gain", "LOH", "Loss", "=")

seg <- GRanges(seqnames=chr,
             ranges=IRanges(start=start, end=end),
             medLAF=medLAF,
             medgLAF=gLAF,
             ratio=ratio,
               somaCN=copynumber,
               event=event)
data(GCcontent)
x <- segmentGCbiasRemoval(seg, data, GCcontent)
```

`segwithratio` *Segments with somatic ratio.*

Description

SomatiCA segmentation and somatic ratio estimation results for data(glio).

Usage

```
data(segwithratio)
```

Format

A data frame with 143 segments on the following 6 variables.

`chromosome` A character vector.

`start` An integer vector.

`end` An integer vector.

`medLAF` A numeric vector. Median LAF in tumor sample.

`medgermlineLAF` A numeric vector. Median LAF in control sample.

`ratio` A numeric vector. Estimated somatic ratio.

SomatiCAFormat-methods

~~ *Methods for Function* SomatiCAFormat *in Package* **SomatiCA** ~~

Description

Convert a data frame or read a file into SomatiCA input format. Remove missing values from input, check validity of data type and convert it into SomatiCA input format, which is a GRanges object.

Methods

signature(data = "character") Filename.

signature(data = "data.frame") Data frame with 7 column (tumor sample with control), including seqnames, start, zygosity, tCount, LAF, tCountN, germLAF.

signature(data = "GRanges") A GRange object.

SomatiCApipe *SomatiCA pipeline.*

Description

First, it uses read depths and lesser allele frequencies (LAF) from mapped short sequence reads to segment the genome and identify candidate CNAs. Second, SomaticCNA estimates the admixture rate from the relative copy-number profile of tumor-normal pair by a Bayesian finite mixture model. Third, SomaticCNA quantifies somatic copy-number and subclonality for each genomic segment to guide its characterization.

Usage

```
SomatiCApipe(input, ncores = 1, collapse.k = 0, method = "mle",
                mcmc = 50000, burnin = 10000, p = 0.001, verbose = TRUE,
                rss = FALSE, adjust=TRUE, threshold1 = 0.01,
                threshold2 = 0.05, S = 0.1, GC, set.admix=NULL, ...)
```

Arguments

input	A GRanges object, usually the output from SomatiCAFormat().
ncores	Number of cores used.
collapse.k	Number of data points collapsed.
method	Method used to estimate somatic ratio of given segment. For the "mle" method somatic ratio is estimated by a maximum likelihood approach. For the "mean" method, somatic ratio is estimated by the ratio between mean of tumor sample and normal sample. For the "geometric", somatic ratio is estimated by geometric mean of somatic ratios of all sites in given segment.

mcmc	Number of iterations used for Markov Chain Monte Carlo.
burnin	Number of iterations of burnin used for Markov Chain Monte Carlo.
p	Posterior probability cutoff for ambiguous segment assignment.
verbose	Whether working messages will be shown.
rss	Whether residue sum of square cutoff will be used.Default is FALSE.
S	Cutoff for the residue sum of square.
adjust	Ajust the read depths to make the median of two libraries equal.
threshold1	The threshold used to merge the segments based on median LAF. Default is 0.01.
threshold2	The threshold used to merge the segments based on somatic ratio. Default is 0.05.
GC	A data frame object with 4 column."chr", "interval1", "interval2" and "GC".
set.admix	A numeric object. Use prefixed admixture rate instead of estimation of SomatiCA. set.admix=NULL
...	Arguments for `larsCBSsegment()`.

Value

admix	Admixture rate.
segment	A GRanges object, segments with annotation of somatic event and subclonality characterization.

Author(s)

Mengjie Chen

Examples

```
rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05),
            rnorm(200, 0.2, 0.05), rnorm(200, 0.3, 0.05), rnorm(250, 0.4, 0.05))
germLAF <- c(rnorm(800+650, 0.4, 0.05))
rawLAF <- ifelse(rawLAF>0.5, 1-rawLAF, rawLAF)
germLAF <- ifelse(germLAF>0.5, 1-germLAF, germLAF)
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60),  rpois(200, 25),
            rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650, 50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(seq(1, 16000000, by=20000), seq(1, 13000000, by=20000))
zygo <- rep("het", 800+650)
x <- data.frame(chr, as.integer(position), as.character(zygo), as.integer(reads1), rawLAF, as.integer(reads2), ge
colnames(x) <- c("seqnames", "start", "zygosity", "tCount", "LAF", "tCountN", "germLAF")
data <- SomatiCAFormat(x)
data(GCcontent)
res <- SomatiCApipe(data, mcmc = 10000, burnin = 5000, rss=FALSE, GC=GCcontent)
```

SomatiCAUsersGuide *View SomatiCA User's Guide*

Description

Finds the location of the SomatiCA User's Guide and opens it.

Usage

```
SomatiCAUsersGuide(view=TRUE)
```

Arguments

view	logical, should the document be opened using the default PDF document reader?

Value

Character string giving the file location. If `view=TRUE`, the PDF document reader is started and the User's Guide is opened, as a side effect.

Author(s)

Mengjie Chen

Examples

```
# To get the location:
SomatiCAUsersGuide(view=FALSE)
# To open in pdf viewer:
## Not run: SomatiCAUsersGuide()
```

somaticRatio *Estimate somatic ratio for given segments.*

Description

Somatic ratio is defined as the ratio of read depths between a tumor and its paired normal sample for a given segment. SomatiCA implements different methods to estimate somatic ratio.

Usage

```
somaticRatio(seg, data, method = "mle", adjust=FALSE)
```

Arguments

seg	A GRanges object, segments from larsCBSsegment().
data	Input data, from SomatiCAFormat().
method	Method used to estimate somatic ratio of given segments. For the "mle" method somatic ratio is estimated by a maximum likelihood approach. For the "mean" method, somatic ratio is estimated by the ratio between mean of tumor sample and normal sample. For the "geometric", somatic ratio is estimated by geometric mean of somatic ratios of all sites in a given segment.
adjust	Adjust the normal and tumor library and make their median equal.

Value

A GRanges object, segments with annotation of estimated somatic ratio.

Author(s)

Mengjie Chen

See Also

See Also larsCBSsegment.

Examples

```
chr <- c("chr1", "chr1", "chr1", "chr2", "chr2", "chr2")
start <- c(1, 300, 600, 1, 200, 400)
end <- c(300, 600, 800, 200, 400, 650)
medLAF <- c(0.2, 0.4, 0.3, 0.2, 0.3, 0.4)
gLAF <- rep(0.4, 6)
seg <- GRanges(seqnames=chr,
                  ranges=IRanges(start=start, end=end),
                  medLAF=medLAF,
                  medgLAF=gLAF)


rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05),
            rnorm(200, 0.2, 0.05), rnorm(200, 0.3, 0.05), rnorm(250, 0.4, 0.05))
germLAF <- c(rnorm(800+650, 0.4, 0.05))
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60),  rpois(200, 25),
            rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650,50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(c(1:800), c(1:650))
zygo <- rep("het", 800+650)
data <- GRanges(seqnames=chr,
                ranges=IRanges(start=position, width=1),
                zygosity=zygo,
                tCount=reads1,
                LAF=rawLAF,
```

```
                  tCountN=reads2,
                  germLAF=germLAF)

x <- somaticRatio(seg, data, method = "mle")
```

---

subclonality *Estimate subclonality for each somatic copy number abberation.*

Description

Subclonality characterization based on hypothesis testing.

Usage

```
subclonality(segment, admix, bin=0.1, sigma=0.1)
```

Arguments

segment	Segments with GC bias corrected read depth ratio. A GRanges object.
admix	Admixture rate of normal cells.
bin	Bin for percentage of subclones.
sigma	Sigma for normal distribution used for testing.

Details

SomatiCA calculates allelic copy number nB and nA in a control sample based on GC corrected read counts. SomatiCA tests whether copy number change in corresponding tumor sample can result in a change of exactly one copy of one allele. If the somatic ratio (corrected by admixture rate) in the corresponding tumor sample is greater than 1, SomatiCA tests for one copy gain , otherwise it tests for one copy loss. With null hypothesis that clonal copy number ratio follows a normal distribution , p-value is calculated for each segment as the probability of obtaining a copy number ratio at least as extreme as the one that was actually observed. Segments with p-value less than 0.05 are classified as subclonal.

Value

A GRanges object, segments with annotation of somatic event and subclonality.

seqnames, start, end, medLAF, ratio, somaCN, event	Same as the output of copynumberCorrected().
clonality	A character vector. Clonality of somatic copy number aberrations, "=", "clonal", "subclonal_gain" or "subclonal_loss".
germCN	An integer vector. Copy number in control sample.
subclonalCN	An integer vector. Aberrated copy number in tumor clones (if it's clonal) or subclones (if it's subclonal).
subpercent	A numeric vector. Percentage of tumor with aberrated copy number.

Author(s)

Mengjie Chen

Examples

```
### generate sequencing input ###
rawLAF <- c(rnorm(300, 0.2, 0.05), rnorm(300, 0.4, 0.05), rnorm(200, 0.3, 0.05), rnorm(200, 0.2, 0.05), rnorm(200,
germLAF <- c(rnorm(800+650, 0.4, 0.05))
reads1 <- c(rpois(300, 25), rpois(300, 50), rpois(200, 60),  rpois(200, 25), rpois(200, 40), rpois(250, 50))
reads2 <- rpois(800+650, 50)
chr <- c(rep("chr1", 800), rep("chr2", 650))
position <- c(seq(1, 16000000, by=20000), seq(1, 13000000, by=20000))
zygo <- rep("het", 800+650)
data <- GRanges(seqnames=chr,
        ranges=IRanges(start=position, width=1),
        zygosity=zygo,
        tCount=reads1,
        LAF=rawLAF,
        tCountN=reads2,
        germLAF=germLAF)

### generate pseudo segments ###

chr <- c("chr1", "chr1", "chr1", "chr2", "chr2", "chr2")
start <- position[c(1, 301, 601, 1, 201, 401)]
end <-  position[c(301, 601, 800, 201, 401, 651)]
medLAF <- c(0.2, 0.4, 0.3, 0.2, 0.3, 0.4)
gLAF <- rep(0.43, 6)
ratio <- c(0.5, 1, 1.3, 0.5, 0.8, 1)
copynumber <- c(1, 2, 3, 1, 3, 2)
event <- c("LOH", "=", "Gain", "LOH", "Loss", "=")

seg <- GRanges(seqnames=chr,
               ranges=IRanges(start=start, end=end),
             medLAF=medLAF,
             medgLAF=gLAF,
             ratio=ratio,
               somaCN=copynumber,
               event=event)
data(GCcontent)
x <- segmentGCbiasRemoval(seg, data, GCcontent)

admix <- 0.2
segmentClonality <- subclonality(x, admix)
```

Index