

Package ‘panelcn.mops’

April 12, 2018

Type Package

Title CNV detection tool for targeted NGS panel data

Version 1.0.0

Date 2017-09-28

Author Verena Haunschmid, Gundula Povysil

Maintainer Gundula Povysil <povysil@bioinf.jku.at>

Description CNV detection tool for targeted NGS panel data.
Extension of the cn.mops package.

License LGPL (>= 2.0)

LazyData TRUE

Imports GenomicRanges, Rsamtools, IRanges, S4Vectors, GenomeInfoDb,
grDevices

Depends R (>= 3.4), cn.mops, methods, utils, stats, graphics

biocViews Sequencing, CopyNumberVariation, CellBiology,
GenomicVariation, VariantDetection, Genetics

RoxygenNote 6.0.1

NeedsCompilation no

Suggests knitr, rmarkdown, RUnit, BiocGenerics

VignetteBuilder knitr

R topics documented:

| | |
|---------------------------------|----|
| control | 2 |
| countBamListInGRanges | 2 |
| countWindows | 3 |
| createResultTable | 3 |
| getWindows | 4 |
| panelcn.mops | 5 |
| plotBoxplot | 6 |
| read.width | 7 |
| resultlist | 7 |
| runPanelcnMops | 8 |
| splitROIs | 9 |
| test | 10 |

| | |
|--------------|-----------|
| Index | 11 |
|--------------|-----------|

| | |
|---------|--|
| control | <i>GRanges object of countWindows with read counts for control samples as elementMetadata.</i> |
|---------|--|

Description

The object was created using the function `countBamListInGRanges` with the enclosed `countWindows` object, a subset of BAM files provided by the 1000 Genomes Project and the `read.width` parameter set to 150.

Details

Control data included in `panelcn.mops`

Author(s)

Gundula Povysil

Examples

```
data(panelcn.mops)
control
```

| | |
|------------------------------------|--|
| <code>countBamListInGRanges</code> | <i>Get read counts for a list of BAM files and given count windows</i> |
|------------------------------------|--|

Description

Get read counts for a list of BAM files and given count windows

Usage

```
countBamListInGRanges(bam.files, countWindows, read.width = 150, ...)
```

Arguments

| | |
|---------------------------|---|
| <code>bam.files</code> | list with absolute or relative paths to BAM files |
| <code>countWindows</code> | data.frame with contents of a BED file as returned by <code>getWindows</code> |
| <code>read.width</code> | <code>read.width</code> parameter for <code>countBamInGRanges</code> or <code>FALSE</code> if actual read width should be extracted from BAM file |
| <code>...</code> | additional parameters |

Value

a `GRanges` object over the `countWindows` with read counts for each sample as `elementMetadata`

Examples

```
bed <- system.file("extdata/Genes_part.bed", package = "panelcn.mops")
countWindows <- getWindows(bed)
## Not run:
testbam <- "SAMPLE1.bam"
test <- countBamListInGRanges(countWindows = countWindows,
                             bam.files = testbam, read.width = 150)

## End(Not run)
```

| | |
|--------------|--|
| countWindows | <i>result object of getWindows - a data.frame with the contents of the provided BED file with an additional gene name and exon name column</i> |
|--------------|--|

Description

Data included in panelcn.mops

Author(s)

Gundula Povysil

Examples

```
data(panelcn.mops)
countWindows
```

| | |
|-------------------|---|
| createResultTable | <i>Creates a user readable result table for the test samples of the genes of interest</i> |
|-------------------|---|

Description

Creates a user readable result table for the test samples of the genes of interest

Usage

```
createResultTable(resultlist, XandCB, countWindows, selectedGenes = NULL,
                  sampleNames)
```

Arguments

| | |
|---------------|--|
| resultlist | result object of runPanelcnMops |
| XandCB | GRanges object of combined read counts of test samples and control samples as returned by getRCRanges or countBamListInGRanges |
| countWindows | data.frame with contents of a BED file as returned by getWindows |
| selectedGenes | vector of names of genes of interest that should be displayed or NULL if all genes are of interest. Default = NULL |
| sampleNames | names of the test samples (basename of the BAM files) |

Value

a data.frame containing the results for the test samples within the genes of interest

Examples

```
data(panelcn.mops)
XandCB <- test
elementMetadata(XandCB) <- cbind(elementMetadata(XandCB),
                                elementMetadata(control))
sampleNames <- colnames(elementMetadata(test))
selectedGenes <- "ATM"
resulttable <- createResultTable(resultlist = resultlist, XandCB = XandCB,
                                countWindows = countWindows,
                                selectedGenes = selectedGenes,
                                sampleNames = sampleNames)
```

getWindows

Convert BED file into data.frame of count windows

Description

Convert BED file into data.frame of count windows

Usage

```
getWindows(filename, chr = FALSE)
```

Arguments

| | |
|----------|---|
| filename | filename of the BED file with absolute or relative path (structure of BED file without header: chromosome, exon start, exon end, exon name) |
| chr | indicates whether naming contains chr prefix |

Value

a data.frame with the contents of the BED file with an additional gene name and exon name column

Examples

```
bed <- list.files(system.file("extdata", package = "panelcn.mops"),
                 pattern = ".bed$", full.names = TRUE)
countWindows <- getWindows(bed)
```

panelcn.mops

*Core copy number detection algorithm for targeted NGS panel data***Description**

This function performs the cn.mops algorithm for copy number detection in NGS data adjusted to targeted NGS panel data including the second quality control.

Usage

```
panelcn.mops(input, testi = 1, geneInd = NULL, classes = c("CN0", "CN1",
  "CN2", "CN3", "CN4"), I = c(0.025, 0.5, 1, 1.5, 2), priorImpact = 1,
  cyc = 20, normType = "quant", sizeFactor = "quant", qu = 0.25,
  quSizeFactor = 0.75, norm = 1, minReadCount = 5, maxControls = 25,
  useMedian = FALSE, returnPosterior = TRUE)
```

Arguments

| | |
|--------------|---|
| input | either an instance of "GRanges" or a raw data matrix, where columns are interpreted as samples and rows as genomic regions. An entry is the read count of a sample in the genomic region. |
| testi | positive integer that gives the index of the test sample in input. Default = 1 |
| geneInd | vector of indices of rows input that are within target genes. These regions are not considered for choosing correlated reference samples. If NULL, all regions are considered for the correlation. Default = NULL |
| classes | vector of characters of the same length as the parameter vector "I". One vector element must be named "CN2". The names reflect the labels of the copy number classes. Default = c("CN0", "CN1", "CN2", "CN3", "CN4"). |
| I | vector of positive real values containing the expected fold change of the copy number classes. Length of this vector must be equal to the length of the "classes" parameter vector. For human copy number polymorphisms the default is c(0.025,0.5,1,1.5,2). |
| priorImpact | positive real value that reflects how strong the prior assumption affects the result. The higher the value the more samples will be assumed to have copy number 2. Default = 1. |
| cyc | positive integer that sets the number of cycles for the algorithm. Usually after less than 15 cycles convergence is reached. Default = 20. |
| normType | type of the normalization technique. Each samples' read counts are scaled such that the total number of reads are comparable across samples. Options are "mean", "median", "poisson", "quant", and "mode". Default = "quant". |
| sizeFactor | parameter for calculating the size factors for normalization. Options are "mean", "median", "quant", and "mode". Default = "quant". |
| qu | Quantile of the normType if normType is set to "quant". Real value between 0 and 1. Default = 0.25. |
| quSizeFactor | Quantile of the sizeFactor if sizeFactor is set to "quant". 0.75 corresponds to "upper quartile normalization". Real value between 0 and 1. Default = 0.75. |
| norm | the normalization strategy to be used. If set to 0 the read counts are not normalized and cn.mops does not model different coverages. If set to 1 the read counts are normalized. If set to 2 the read counts are not normalized and cn.mops models different coverages. Default = 1. |

| | |
|-----------------|---|
| minReadCount | if all samples are below this value the algorithm will return the prior knowledge. This prevents that the algorithm from being applied to segments with very low coverage. Default = 5. |
| maxControls | integer reflecting the maximal numbers of controls to use. If set to 0 all highly correlated controls are used. Default = 25 |
| useMedian | flag indicating whether "median" instead of "mean" of a segment should be used for the CNV call. Default = FALSE. |
| returnPosterior | flag that decides whether the posterior probabilities should be returned. The posterior probabilities have a dimension of samples times copy number states times genomic regions and therefore consume a lot of memory. Default = TRUE. |

Value

an instance of "CNVDetectionResult".

Examples

```
data(panelcn.mops)
XandCB <- test
elementMetadata(XandCB) <- cbind(elementMetadata(XandCB),
                                elementMetadata(control))
result <- panelcn.mops(XandCB)
```

| | |
|-------------|--|
| plotBoxplot | <i>Create box plot of normalized read counts</i> |
|-------------|--|

Description

Create box plot of normalized read counts

Usage

```
plotBoxplot(result, sampleName, countWindows, selectedGenes = NULL,
            showGene = 1, showLegend = TRUE, exonRange = NULL, ylimup = 1.15,
            thresh = 0)
```

Arguments

| | |
|---------------|--|
| result | result object of panelcn.mops |
| sampleName | name of the test sample that should be displayed |
| countWindows | data.frame with contents of a BED file as returned by getWindow |
| selectedGenes | vector of names of genes of interest that should be displayed or NULL if all genes are of interest. Default = NULL |
| showGene | integer indicating which of the genes of interest to plot |
| showLegend | flag to indicate whether to display a legend with the names of the test samples. Default = TRUE |
| exonRange | vector of 2 positive integers to limit box plot to a certain range of exons or NULL |

ylimup numeric, maximum RC is multiplied by this value to calculate second value of ylim. Default = 1.15

thresh numeric threshold for plotting fold change areas E.g. thresh = 0.4 plots a green rectangle above $(1 + 0.4) * \text{median}$ for each boxplot and a red rectangle below $(1 - 0.4) * \text{median}$. Default of zero does not plot any colored areas.

Value

generates a boxplot of the normalized read counts

Examples

```
data(panelcn.mops)
sampleNames <- colnames(elementMetadata(test))
selectedGenes <- "ATM"
plotBoxplot(result = resultlist[[1]], sampleName = sampleNames[1],
            countWindows = countWindows, selectedGenes = selectedGenes,
            showGene = 1)
```

| | |
|------------|--|
| read.width | <i>read width used for calculating RCs of test and control</i> |
|------------|--|

Description

Data included in panelcn.mops

Author(s)

Gundula Povysil

Examples

```
data(panelcn.mops)
read.width
```

| | |
|------------|--|
| resultlist | <i>result object of runPanelcnMops - a list of instances of "CNVDetectionResult"</i> |
|------------|--|

Description

Result data included in panelcn.mops

Author(s)

Gundula Povysil

Examples

```
data(panelcn.mops)
resultlist
```

| | |
|----------------|--|
| runPanelcnMops | <i>Full copy number detection for targeted NGS panel data for multiple samples</i> |
|----------------|--|

Description

This function performs first quality control and runs panelcn.mops for CNV detection on all test samples.

Usage

```
runPanelcnMops(XandCB, testiv = c(1), countWindows, selectedGenes = NULL,
  I = c(0.025, 0.57, 1, 1.46, 2), normType = "quant",
  sizeFactor = "quant", qu = 0.25, quSizeFactor = 0.75, norm = 1,
  priorImpact = 1, minMedianRC = 30, maxControls = 25, sex = "mixed")
```

Arguments

| | |
|---------------|---|
| XandCB | GRanges object of combined read counts of test samples and control samples as returned by countBamListInGRanges |
| testiv | vector of indices of test samples in XandCB. Default = c(1) |
| countWindows | data.frame with contents of a BED file as returned by getWindow |
| selectedGenes | vector of names of genes of interest or NULL if all genes are of interest. Default = NULL |
| I | vector of positive real values containing the expected fold change of the copy number classes. Length of this vector must be equal to the length of the "classes" parameter vector. For targeted NGS panel data the default is c(0.025,0.57,1,1.46,2) |
| normType | type of the normalization technique. Each samples' read counts are scaled such that the total number of reads are comparable across samples. Options are "mean", "median", "poisson", "quant", and "mode" Default = "quant" |
| sizeFactor | parameter for calculating the size factors for normalization. Options are "mean", "median", "quant", and "mode". Default = "quant" |
| qu | Quantile of the normType if normType is set to "quant". Real value between 0 and 1. Default = 0.25 |
| quSizeFactor | Quantile of the sizeFactor if sizeFactor is set to "quant". 0.75 corresponds to "upper quartile normalization". Real value between 0 and 1. Default = 0.75 |
| norm | the normalization strategy to be used. If set to 0 the read counts are not normalized and cn.mops does not model different coverages. If set to 1 the read counts are normalized. If set to 2 the read counts are not normalized and cn.mops models different coverages. Default = 1. |
| priorImpact | positive real value that reflects how strong the prior assumption affects the result. The higher the value the more samples will be assumed to have copy number 2. Default = 1 |
| minMedianRC | segments with median read counts over all samples < minMedianRC are excluded from the analysis |
| maxControls | integer reflecting the maximal numbers of controls to use. If set to 0 all highly correlated controls are used. Default = 25 |
| sex | either "mixed", "male", or "female" reflecting the sex of all samples (test and control) |

Value

list of instances of "CNVDetectionResult"

Examples

```
data(panelcn.mops)
XandCB <- test
elementMetadata(XandCB) <- cbind(elementMetadata(XandCB),
                                  elementMetadata(control))
resultlist <- runPanelcnMops(XandCB, countWindows = countWindows)
```

| | |
|-----------|---|
| splitROIs | <i>Split (larger) ROIs into multiple smaller (overlapping) bins and create new BED file</i> |
|-----------|---|

Description

Split (larger) ROIs into multiple smaller (overlapping) bins and create new BED file

Usage

```
splitROIs(oldBedFile, newBedFile, limit = 0, bin = 100, shift = 50,
           chr = FALSE)
```

Arguments

| | |
|------------|---|
| oldBedFile | filename of the BED file with absolute or relative path (structure of BED file without header: chromosome, exon start, exon end, exon name) |
| newBedFile | filename of the new BED file that should be created |
| limit | ROIs larger than limit will be split |
| bin | size of bins (in bp) the ROIs will be split into |
| shift | no. of bp between start positions of adjacent bins |
| chr | indicates whether naming contains chr prefix |

Value

generates a new BED file with (larger) ROIs split into smaller bins

Examples

```
bed <- list.files(system.file("extdata", package = "panelcn.mops"),
                 pattern = ".bed$", full.names = TRUE)
splitROIs(bed, "newBed.bed")
```

`test`*GRanges object of countWindows with read counts for a test sample as elementMetadata.*

Description

The object was created using the function `countBamListInGRanges` with the enclosed `countWindows` object, a subset of a BAM file provided by the 1000 Genomes Project and the `read.width` parameter set to 150.

Details

Test data included in `panelcn.mops`

Author(s)

Gundula Povysil

Examples

```
data(panelcn.mops)
test
```

Index

*Topic **data**

- control, [2](#)
- countWindows, [3](#)
- read.width, [7](#)
- resultlist, [7](#)
- test, [10](#)

- control, [2](#)
- countBamListInGRanges, [2](#)
- countWindows, [3](#)
- createResultTable, [3](#)

- getWindows, [4](#)

- panelcn.mops, [5](#)
- plotBoxplot, [6](#)

- read.width, [7](#)
- resultlist, [7](#)
- runPanelcnMops, [8](#)

- splitROIs, [9](#)

- test, [10](#)