Package 'CopyNumber450kCancer'

June 18, 2015

Type Package

Title Baseline Correction for Copy Number Data from Cancer Samples

Version 1.0.3

Date 2015-06-17

Author Nour-al-dain Marzouka [aut, cre]

Maintainer Nour-al-dain Marzouka <nour.marzouka@medsci.uu.se>

Description The 450k arrays are frequently used in the epigenetic studies, the copy number calling from the 450k data is possible but it faces some difficulties in cancer samples regarding the determination of the copy number status due to the false sample centering and baseline shifting. Without solving this issue the CN calling will be inaccurate. CopyNumber450kCancer-package was designed to correct the baseline in cancer samples using the Maximum Density Peak Estimation (MDPE) method.
The main advantages for CopyNumber450kCancer-package are: Fast (few seconds per sample), high accuracy rate, in-sample correction, no input parameters needed, low computer sources required, and adaptable for 450k-similar technologies.

Depends R (>= 3.1.0)

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2015-06-18 13:38:49

R topics documented:

CopyNumber450kCancer-package
Baseline Correction for Accurate Copy Number Calling in 450k Methylation Array

---

Description

The 450k arrays are frequently used in the epigenetic studies, the copy number calling from the 450k data is possible but it faces some difficulties in cancer samples regarding the determination of the copy number status due to the false sample centering and baseline shifting. Without solving this issue the copy number calling will be inaccurate. CopyNumber450kCancer-package was designed to correct the baseline in cancer samples using the Maximum Density Peak Estimation. The main advantages for CopyNumber450kCancer-package are: Fast (few seconds per sample), high accuracy rate, in-sample correction, no input parameters needed, low computer sources required, and adaptable for 450k-similar technologies.

Details

Package:	CopyNumber450kCancer
Type:	Package
Version:	1.0.3
Date:	2015-06-17
License:	GPL (>= 2)

Author(s)

Nour-al-dain Marzouka

Maintainer: Nour-al-dain Marzouka <nour.marzouka@medsci.uu.se>

See Also

https://github.com/nourmarzouka/CopyNumber450kCancer

AutoCorrectPeak
This Function is the Main Function in CopyNumber450kCancer Package Which Correct the Baseline Based on the Density of the Probes/Segments

---

Description

This function generate: 1) a new segmentation file which is similar to the segmentation input but with the new log values, 2) plots show the samples before and after the baseline correction, 3) a QC file, 4) a file contains the shifting amounts for each sample

QC: There are no well-defined quality control (QC) standards for 450k data segmentation; therefore we selected different quality control standards those already used with the SNP arrays. QC file will be produced with these QC standards for each sample: Interquartile Range (IQR), Median Absolute Pairwise Difference (MAPD), and number of the segments, standard deviation (SD), and in the addition of a suggested QC standard named Maximum Density Peak Sharpness (MDPS). The QC values are calculated based on the log values of the segments but not the probes. BaCo450k does not provide any QC thresholds because they are different from an experiment to another. The user can check the QC file and exclude the samples those have low quality QC values. We strongly recommend the visual reviewing for the segmentation plots as to recognize the low-quality samples.

Usage

```
AutoCorrectPeak(object, cutoff = 0.1, markers = 20, ...)
```

Arguments

object	Data object, as returned by [ReadData]().
cutoff	The cutoff to be used in the ploting, the data or the new segments file will not be affected (default: 0.1)
markers	The minimum number of the probes required in the segment to be plotted (default: 20)
...	Sent to [plotRegions]().

Examples

```
#example
# the package contains example files: regions.csv and sample_list.csv
#to load the example regions.csv and sample_list.csv files
regions <- system.file("extdata", "regions.csv", package="CopyNumber450kCancer")
sample_list <- system.file("extdata", "sample_list.csv", package="CopyNumber450kCancer")

# Creat the object for the package
object <- ReadData(regions,sample_list)

# Baseline autocorrection,
# this will creat different plot and QC and new regions file in the working directory
object <- AutoCorrectPeak(object)

# For manual revision and manual baseline determination
# object <- ReviewPlot(object)

# To plot the final plots
PlotCNV(object) # to plot all the samples
PlotCNV(object, select= c(1,4), comment=FALSE, cutoff=0.1, markers=20) # to plot some samples
```

PlotCNV   *Plot the CNV Regions with the Density Plot*

Description

Uses the regions object, this function will plot the last modified regions object (i.e. after the auto-correction or the reviewing)

Usage

```
PlotCNV(object, select, comment = F, cutoff = 0.1, markers = 20)
```

Arguments

object	Data object, as returned by [ReadData](#).
select	The number of the samples to be reviewed, if not determined then all the samples in the sample list will be selected
comment	If TRUE then the comment in the sample list file will be printed in the header of the plots, default is FALSE
cutoff	cutoff for the segments coloring , default is cutoff = 0.1
markers	the minimum number of markers in the segment to be plotted, default is markers = 20

Examples

```
#example
# the package contains example files: regions.csv and sample_list.csv
#to load the example regions.csv and sample_list.csv files
regions <- system.file("extdata", "regions.csv", package="CopyNumber450kCancer")
sample_list <- system.file("extdata", "sample_list.csv", package="CopyNumber450kCancer")

# Creat the object for the package
object <- ReadData(regions,sample_list)

# Baseline autocorrection,
# this will creat different plot and QC and new regions file in the working directory
object <- AutoCorrectPeak(object)

# For manual revision and manual baseline determination
# object <- ReviewPlot(object)

# To plot the final plots
PlotCNV(object) # to plot all the samples
PlotCNV(object, select= c(1,4), comment=FALSE, cutoff=0.1, markers=20) # to plot some samples
```

PlotMerged   *To Make Plots with Merged Regions*

Description

optional function to make all the regions in 3 fixed log levels (deletion, normal, duplication) and then to plot the samples

Usage

```
PlotMerged(object, cutoff = 0.1, markers = 20, ...)
```

Arguments

object	Data object, as returned by [ReadData].
cutoff	This cutt of will affect these plots, default is cutoff = 0.1
markers	minimum number of marker required in the region, default is (markers = 20)
...	

Examples

```
# the package contains example files: regions.csv and sample_list.csv
#to load the example regions.csv and sample_list.csv files
regions <- system.file("extdata", "regions.csv", package="CopyNumber450kCancer")
sample_list <- system.file("extdata", "sample_list.csv", package="CopyNumber450kCancer")

# Creat the object for the package
object <- ReadData(regions,sample_list)

# Optional function to plot all the amplifications and delections in the same level
PlotMerged(object,comment = TRUE, cutoff = 0.2, markers = 5)
```

---

plotRegions   *Genomic Copy Number Plotting Function*

Description

To be used internally by the package only.

This function plots the chromosomal regions (segments) with colored segments based on the cutoff.
This function was built based on "plotSample" function in "CopyNumber450k" package (http://www.bioconductor.org/packag
and uses a modified "minor.tick" function in "Hmisc" package to draw small tick in the plots.

Usage

```
plotRegions(object, chr, start, end, cutoff = 0.1, markers = 20, ...)
```

Arguments

`object`	Data object, as returned by [ReadData]().
`chr`	Chromosomes to plot.
`start`	Start positions of region to plot.
`end`	End position of region to plot.
`cutoff`	Log R ratio cutoff.
`markers`	Minimum number of markers per segment cutoff. Less than this number the segment will not be plotted.
`...`	Sent to [plot]().

Examples

```
#To be used internally by the package only

# the package contains example files: regions.csv and sample_list.csv
#to load the example regions.csv and sample_list.csv files
regions <- system.file("extdata", "regions.csv", package="CopyNumber450kCancer")
sample_list <- system.file("extdata", "sample_list.csv", package="CopyNumber450kCancer")

# Creat the object for the package
object <- ReadData(regions,sample_list)

#to plot all the regions and all the sampls in one plot
plotRegions(object$regions)
```

```
print.CopyNumber450kCancer_data
```
Print Function for CopyNumber450kCancer Data(x)

Description

Print function.

Arguments

`x`	Data object, as returned by [ReadData]().
`...`	Ignored.

ReadData	*Function Reads the Data (i.e. regions file and sample list file) for CopyNumber450kCancer*

Description

The input should be two files regions file and sample list file. regions file: contains the data for all the regions/segments in all the sample sample list file: contains the number of the samples, the names of the samples and user comment.

The header of the segments/regions file should be in this order and with these names: "Sample", "Chromosome", "bp.Start", "bp.End", "Num.of.Markers", "Mean" The segments file should have all the samples in one file Be carful for the dots and it is case sensitive. Sample: the sample name Chromosome: the chromosome number chr1, chr2, ..., chrX, chrY bp.Start: number, the start point for the segment bp.End: number, the end point for the segment Num.of.Markers: Mean: is the log value for the segment

The header of the sample list file should be in this order and with these names: To check if the header of the sample list file is ok "Number", "Sample", "Comment" Be carful it is case sensitive. Number: is the number of the sample 1,2,3,.... Sample: the name of the samples Comment: any comment the user want to see in the reviewing step and in the QC file, ex karyotyping.

Usage

```
ReadData(regions_file, Sample_list, copynumber450k = F)
```

Arguments

regions_file   The segmentaion file (CSV file)

Sample_list   The CSV file that contains the names of the samples and the user comments

copynumber450k True if the file is the output of copynumber450k, defualt is FALSE.

Examples

```
#example
# the package contains example files: regions.csv and sample_list.csv
#to load the example regions.csv and sample_list.csv files
regions <- system.file("extdata", "regions.csv", package="CopyNumber450kCancer")
sample_list <- system.file("extdata", "sample_list.csv", package="CopyNumber450kCancer")

# Creat the object for the package
object <- ReadData(regions,sample_list)

# Baseline autocorrection,
# this will creat different plot and QC and new regions file in the working directory
object <- AutoCorrectPeak(object)

# For manual revision and manual baseline determination
# object <- ReviewPlot(object)
```

```
# To plot the final plots
PlotCNV(object) # to plot all the samples
PlotCNV(object, select= c(1,4), comment=FALSE, cutoff=0.1, markers=20) # to plot some samples
```

ReviewPlot   *Interactive Plots for Baseline Correction in Copy Number Data*

Description

CopyNumber450kCancer provides graphical interactive plots as an option to supervise/review the
baseline estimation. The user can click on the LRR levels where the baseline should be located in
between. The segments in the defined range will be selected and the user will be asked to choose
one of two options for how the new baseline should be calculated. The first option is the weighted
median for the LRRs with the respect of the number of the probes each segment. The second
option is the local highest peak in the probes density function. For the reviewing step, we strongly
recommend to use any other information (ex. karyotyping) those can help the user in deciding.

This function uses modified "weighted.median" function from isotone package.

Usage

```
ReviewPlot(object, select, plots = T, cutoff = 0.1, markers = 20, ...)
```

Arguments

object	Data object, as returned by [ReadData](#).
select	The number of the samples the you want to review of modify the baseline for them, if not specified then all the samples in the sample list will be selected
plots	To get plots after the revieing all the selected samples (default is TURE) if FALSE then you will get only the new segments file (reviewed_regions) and a file contains the recording of the reviewing process.
cutoff	The cutoff to be used in the ploting only, the data or the new segments file will not be affected (default: 0.1)
markers	The minimum number of the probes required in the segment to be plotted (default: 20)
...	

Examples

```
#example
# the package contains example files: regions.csv and sample_list.csv
#to load the example regions.csv and sample_list.csv files
regions <- system.file("extdata", "regions.csv", package="CopyNumber450kCancer")
sample_list <- system.file("extdata", "sample_list.csv", package="CopyNumber450kCancer")

# Creat the object for the package
```

```
object <- ReadData(regions,sample_list)

# Baseline autocorrection,
# this will creat different plot and QC and new regions file in the working directory
object <- AutoCorrectPeak(object)

# For manual revision and manual baseline determination
# object <- ReviewPlot(object)

# To plot the final plots
PlotCNV(object) # to plot all the samples
PlotCNV(object, select= c(1,4), comment=FALSE, cutoff=0.1, markers=20) # to plot some samples
```

Index