

KCsmart

April 19, 2009

KCData-class

Internal class "KCData"

Description

Internal superclass "KCData"

Objects from the Class

The user is not meant to create instances of this class

Slots

data: Internal data

Methods

[[< signature(x = "KCData"): ...

[[signature(x = "KCData"): ...

length signature(x = "KCData"): ...

unlist signature(x = "KCData"): ...

Warning

This class is meant for internal use only

Note

Internal class

Author(s)

Jorma de Ronde

KCsmart-package *KCsmart*

Description

Multiple sample aCGH analysis using kernel convolution

Details

Package: KCsmart
Type: Package
Version: 1.0
Date: 2008-08-14
License: GPL

Use the wrapper function 'calcSpm' to calculate the sample point matrix. Use 'findSigLevelTrad' to find a significance threshold using permutation based testing. Use 'plot' to plot the sample point matrix or 'plotScaleSpace' to plot the significant regions over multiple scales (sigmas). Use 'getSigSegments' to retrieve the significantly gained and lost regions using specific cutoffs.

Author(s)

Jorma de Ronde, Christiaan Klijn
Maintainer: Jorma de Ronde <j.d.ronde@nki.nl>

References

Identification of cancer genes using a statistical framework for multiexperiment analysis of nondiscretized array CGH data. Nucleic Acids Res. 2008 Feb;36(2):e13.

See Also

[calcSpm](#), [findSigLevelTrad](#), [findSigLevelFdr](#), [plot](#), [plotScaleSpace](#), [getSigSegments](#)

Examples

```
data (hsSampleData)
data (hsMirrorLocs)

spm1mb <- calcSpm (hsSampleData, hsMirrorLocs)
spm4mb <- calcSpm (hsSampleData, hsMirrorLocs, sigma=4000000)

plot (spm1mb)
plot (spm1mb, chromosomes=c(1,5,6,'X'))

siglevel1mb <- findSigLevelTrad (hsSampleData, spm1mb, n=3)
siglevel4mb <- findSigLevelTrad (hsSampleData, spm4mb, n=3)

plot (spm1mb, sig.level=siglevel1mb)

plotScaleSpace (list (spm1mb, spm4mb), list (siglevel1mb, siglevel4mb), type='g')
```

```
sigSegments1mb <- getSigSegments(spm1mb, siglevel1mb)
```

KcghData-class *Class "KcghData"*

Description

Internal class

Slots

probeAnnotation: Object of class "probeAnnotation"

data: Holds aCGH data

Methods

initialize signature(.Object = "KcghData"): Internal use only

sort signature(x = "KcghData", decreasing = "missing"): Internal use only

Note

For internal use only

Author(s)

Jorma de Ronde

KcghDataMirror-class
Class "KcghDataMirror"

Description

Internal class

Slots

mirrorLocs: Holds mirrorLocs object

probeAnnotation: Object of class "probeAnnotation"

pos: Holds aCGH data for losses

neg: Holds aCGH data for gains

nrSamples: The number of samples in this analysis

Extends

Class "[KcghDataSum](#)", directly.

Methods

initialize signature(.Object = "KcghDataMirror"): For internal use only

Note

For internal use only

Author(s)

Jorma de Ronde

KcghDataSplit-class

Class "KcghDataSplit"

Description

Internal class

Slots

probeAnnotation: Object of class "probeAnnotation"

pos: Holds aCGH data for losses

neg: Holds aCGH data for gains

Methods

initialize signature(.Object = "KcghDataSplit"): Internal use only

Note

For internal use only

Author(s)

Jorma de Ronde

KcghDataSum-class *Class "KcghDataSum"*

Description

Internal class

Slots

probeAnnotation: Object of class "probeAnnotation"

pos: Holds aCGH data for losses

neg: Holds aCGH data for gains

nrSamples: The number of samples in this analysis

Methods

initialize signature(.Object = "KcghDataSum"): For internal use only

sort signature(x = "KcghDataSum", decreasing = "missing"): For internal use only

Note

For internal use only

Author(s)

Jorma de Ronde

calcSpm

KCsmart wrapper

Description

Wrapper function that calculates the sample point matrix from the aCGH data

Usage

```
calcSpm(data, mirrorLocs, sigma = 1e+06, sampleDensity = 50000, maxmem = 1000)
```

Arguments

data	The aCGH data. Can either be in DNACopy format or as a data.frame described in the details section
mirrorLocs	List containing the chromosome start, centromere and end positions
sigma	The kernel width
sampleDensity	The sample point matrix resolution
maxmem	This parameter controls memory usage, set to lower value to lower memory consumption

Details

'data' can be in cghRaw (CGHbase), DNACopy or in data.frame format. When using the latter, the data.frame must have the following two columns: 'chrom' stating the chromosome the probe is located on, 'maploc' describing the position on the chromosome of the probe. The remainder of the data.frame will be interpreted as sample data points. The row names of that data will be used as probe names (when available). Important note: the data can not contain any missing values. If your data includes missing values you will need to preprocess (for example impute) it using other software solutions.

The mirror locations for Homo Sapiens and Mus Musculus are provided in the package. These can be loaded using data(hsMirrorLocs) and data(mmMirrorLocs) respectively. The 'mirrorLocs' object is a list with vectors containing the start, centromere (optional) and end of each chromosome as the list elements. Additionally it should contain an attribute 'chromNames' listing the chromosome names of each respective list element.

'sigma' defines the kernel width of the kernel used to convolute the data.

'sampleDensity' defines the resolution of the sample point matrix to be calculated. A sampleDensity of 50000 would correspond to a sample point every 50k base pairs.

Value

Returns a sample point matrix object. The object has several slots of which the 'data' slot contains a list where each list item represents a chromosome. Each list item in turn contains the sample point matrix for the gains and the losses separately and an attribute specifying the corresponding chromosome. The sample point matrix contains the following additional slots: totalLength: Total length of the sample point matrix maxy and miny: Maximal and minimal score attained

The other slots just represent the parameters used to calculate the sample point matrix.

Use 'plot' to plot the sample point matrix and 'findSigLevelTrad' to find a significance threshold. 'plotScaleSpace' can be used to plot the significant regions of multiple sample point matrices (using different sigmas).

Author(s)

Jorma de Ronde

See Also

[plot](#), [findSigLevelTrad](#), [plotScaleSpace](#)

Examples

```
data(hsSampleData)
data(hsMirrorLocs)

spm1mb <- calcSpm(hsSampleData, hsMirrorLocs)
spm4mb <- calcSpm(hsSampleData, hsMirrorLocs, sigma=4000000)

plot(spm1mb)
plot(spm1mb, chromosomes=c(1,5,6,'X'))
```

findSigLevelFdr *This function has not been properly implemented yet*

Description

Method to find the cutoff at which gains and losses are considered significant using permutations

Usage

```
findSigLevelFdr(data, observedSpm, n = 1, fdrTarget=0.05, maxmem=1000)
```

Arguments

data	aCGH data in the same format as used for 'calcSpm'
observedSpm	A sample point matrix as produced by 'calcSpm'
n	Number of permutations
fdrTarget	Target False Discovery Rate (FDR)
maxmem	This parameter controls memory usage, set to lower value to lower memory consumption

Details

The number of permutations needed for reliable results depends on the data and can not be determined beforehand. As a general rule-of-thumb around 100 permutations should be used for 'quick checks' and around 2000 permutations for more rigorous testing. The FDR method is less conservative than the p-value based approach since instead of controlling the family wise error rate (FWER, $P(\text{false positive} > 1)$) it controls the false discovery rate (FDR) (false positives / total number of called data points).

Value

A list with the cutoffs corresponding to the given FDR

pos	The cutoff for the gains
neg	The cutoff for the losses'

Author(s)

Jorma de Ronde

See Also

[plotScaleSpace](#)

Examples

```

data (hsSampleData)
data (hsMirrorLocs)

spm1mb <- calcSpm(hsSampleData, hsMirrorLocs)

sigLevel1mb <- findSigLevelTrad(hsSampleData, spm1mb, n=3)

plot (spm1mb, sigLevels=sigLevel1mb)
plotScaleSpace(list (spm1mb), list (sigLevel1mb), type='g')

```

findSigLevelTrad *Find significance level*

Description

Method to find the cutoff at which gains and losses are considered significant using permutations

Usage

```
findSigLevelTrad(data, observedSpm, n = 1, p = 0.05, maxmem = 1000)
```

Arguments

data	aCGH data in the same format as used for 'calcSpm'
observedSpm	A sample point matrix as produced by 'calcSpm'
n	Number of permutations
p	Alpha level for significance
maxmem	This parameter controls memory usage, set to lower value to lower memory consumption

Details

The number of permutations needed for reliable results depends on the data and can not be determined beforehand. As a general rule-of-thumb around 100 permutations should be used for 'quick checks' and around 2000 permutations for more rigorous testing.

p is the uncorrected alpha level, the method corrects for multiple testing internally using simple Bonferroni correction. See the referenced publication for more details.

Value

A list with the cutoffs corresponding to the given alpha level

pos	The cutoff for the gains
neg	The cutoff for the losses'

Author(s)

Jorma de Ronde

See Also[plotScaleSpace](#)**Examples**

```
data (hsSampleData)
data (hsMirrorLocs)

spm1mb <- calcSpm(hsSampleData, hsMirrorLocs)

sigLevel1mb <- findSigLevelTrad(hsSampleData, spm1mb, n=3)

plot (spm1mb, sigLevels=sigLevel1mb)
plotScaleSpace(list (spm1mb), list (sigLevel1mb), type='g')
```

getSigSegments	<i>Retrieve the significantly gained and lost regions including the corresponding, original probes</i>
----------------	--

Description

Retrieve the significantly gained and lost regions including the corresponding, original probes. A significance level must be selected by the user.

Usage

```
getSigSegments(spm, sigLevels, chromosomes=NULL)
```

Arguments

spm	The sample point matrix to be plotted
sigLevels	The significance thresholds to be used
chromosomes	Takes a vector of chromosomes to be plotted. Defaults to all chromosomes.

Details

'sigLevels' should contain the significance thresholds in a list with the positive (gains) threshold in the 'pos' element and the negative (losses) threshold in the 'neg' element. This is the format as returned by 'findSigLevelTrad' and 'findSigLevelFdr'.

Value

Returns a sigSegments object containing the chromosome, start position, end position, average KC score and the mode of the KC score in that region of all segments passing the thresholds as set in 'sigLevels'. Additionally, returns the IDs and indices of the probes and the positions in the sample point matrix within the significant regions. The results are stored in two separate slots: 'gains' for gains and 'losses' for losses. Use 'write.table' to save the results to file.

Author(s)

Jorma de Ronde

References

~put references to the literature/web site here ~

See Also

[findSigLevelTrad](#), [findSigLevelTrad](#), [write.table](#)

Examples

```
data(hsSampleData)
data(hsMirrorLocs)

spm1mb <- calcSpm(hsSampleData, hsMirrorLocs)

siglevel1mb <- findSigLevelTrad(hsSampleData, spm1mb, n=3)

sigSegments1mb <- getSigSegments(spm1mb, siglevel1mb)
write.table(sigSegments1mb, file=file.path(tempdir(), 'sigSegments1mb.txt'))
```

hsMirrorLocs

Mirror locations of the human genome

Description

Mirror locations of the human genome, based on the NCBI 36 assembly of the human genome, for use with the KCsmart package.

Usage

```
hsMirrorLocs
```

Format

A list containing for each chromosome the start and end position and the centromere location (if a centromere is present).

Source

Ensembl

References

<http://www.ensembl.org>

hsSampleData	<i>Homo Sapiens artificial cgh data set</i>
--------------	---

Description

An artificial cgh data set, created by permuting a BAC data set consisting of 20 samples and introducing an artificial gain on 1p. To be used with the KCsmart package.

Usage

```
hsSampleData
```

Format

A data.frame containing 3268 rows and 22 columns

Source

Artificial data set

idPoints	<i>Identify points in sample point matrix plot</i>
----------	--

Description

Identify points in sample point matrix plot

Usage

```
idPoints(spm, mode='pos', dev=2, chromosomes=NULL)
```

Arguments

spm	The sample point matrix object of which points are to be identified
mode	Determines which points will be identified: mode='pos' will identify points in gained regions, mode='neg' will identify points in lost regions
dev	The device on which the sample point matrix was plotted
chromosomes	If not all chromosomes contained in the sample point matrix were plotted (using the 'chromosomes' argument in the 'plot' command), the same chromosomes must be entered here as an argument

Details

Using the mouse pointer points in a sample point matrix plot can be identified by left-clicking on the to-be-identified points. Right-clicking exits the selection and returns the selected points.

Value

Returns a data.frame listing the the position and the KC score for each identified point.

KCscore	KCscore of the identified point
chromosome	Chromosome on which the identified point is located
chromPosition	Position on the chromosome of the identified point
colin	Co-linear location of the identified point (given the selected chromosomes)

Author(s)

Jorma de Ronde

See Also

[plot](#)

Examples

```
data (hsSampleData)
data (hsMirrorLocs)

#spm1mb <- calcSpm (hsSampleData, hsMirrorLocs)

#plot (spm1mb, type=1)
#idPoints (spm1mb)

#x11 ()
#plot (spm1mb, chromosomes=c (1,2,5))
#idPoints (spm1mb, mode='neg', dev=3, chromosomes=c (1,2,5))
```

mmMirrorLocs

Mirror locations of the mouse genome

Description

Mirror locations of the mouse genome, based on the NCBI m37 mouse assembly, for use with the KCsmart package.

Usage

```
mmMirrorLocs
```

Format

A list containing for each chromosome the start and end position.

Source

Ensembl

References

<http://www.ensembl.org>

plot

*Plot a sample point matrix***Description**

Plot the sample point matrix or parts of it

Usage

```
plot(x,y, ...)
## S4 method for signature 'scaleSpace, missing':
plot(x, y, spm, type='b', ...)
## S4 method for signature 'samplePointMatrix, missing':
plot(x, y, type="b", sigLevels=NULL, chromosomes=NULL, colinAxis=NULL, fillColor
```

Arguments

x	either an object of class samplePointMatrix or scaleSpace
y	object of class missing
type	Determines which data is plotted. 'g' for gains only, 'l' for losses only and 'b' and '1' for both in one plot device
spm	add stuff here
sigLevels	If given, the cutoffs will be drawn as lines in the plots. Optional
chromosomes	Takes a vector of chromosomes to be plotted. Defaults to all chromosomes.
colinAxis	Allows you to override default behaviour of axis labeling. Choose False for genomic position labeling for each individual chromosome, True for colinear labeling.
fillColor	Allows you to choose the colors used to fill the significant areas under the curve. Takes a list with the 'pos' element giving the color for the gains and the 'neg' element the color for the losses.
maploc	Currently not in use
interpolation	Determines which points from the sample point matrix will actually be plotted. If the value of 'interpolation' is n, then every n-th point will be plotted. The default value of 1 will result in all points being plotted. This can be useful when a high density sample point matrix results in big file size when exporting the image (especially to pdf or eps format).
...	Any other parameters you would like to pass to 'plot'. See 'par' for more details.

Value

Plots the sample point matrix. The gains and the losses are plotted separately. The KC normalized score is plotted on the y-axis, the genomic position on the x-axis. If centromeres are present these are represented by dotted, lightblue lines. Setting type to 'b' or to '1' will both make the plot appear in one plot device, '1' will plot the gains and the losses in one plot, 'b' will plot the gains and losses separately. The function 'idPoints' can be used to identify points in the sample point matrix plot. See the corresponding documentation for details.

Author(s)

Jorma de Ronde

See Also[calcSpm](#), [plotScaleSpace](#), [idPoints](#)**Examples**

```

data(hsSampleData)
data(hsMirrorLocs)

spm1mb <- calcSpm(hsSampleData, hsMirrorLocs)

plot(spm1mb)
plot(spm1mb, interpolation=10)
plot(spm1mb, chromosomes=c(1,4,'X'))

siglevel1mb <- findSigLevelTrad(hsSampleData, spm1mb, n=3)
plot(spm1mb, chromosomes=c(1,4,'X'), sigLevels=siglevel1mb)
plot(spm1mb, chromosomes=c(1,4,'X'), sigLevels=siglevel1mb, fillColor=list(pos='darkred',

```

plotScaleSpace

*Plot multiple significant regions in one figure***Description**

Plots significant regions in different scale spaces in one figure

Usage

```
plotScaleSpace(spms, sigLevels, chromosomes=NULL, type='b')
```

Arguments

spms	List of sample point matrices
sigLevels	List of significance levels
chromosomes	Takes a vector of chromosomes to be plotted. Defaults to all chromosomes.
type	Determines which data is plotted. 'g' for gains only, 'l' for losses only and 'b' for both. When type='b' is used, two devices (x11) will be opened.

Details

Takes sample point matrices that were calculated using (different) kernel widths (sigma), then calculates the significant regions given the cutoffs as defined by 'sigLevels' and plots these in one figure.

Value

Depending on the 'type' parameter, produces one or two plots, one for the gains and one for the losses. The heatmap color indicates the level of the gain or loss.

Author(s)

Jorma de Ronde

See Also[plot](#)**Examples**

```
data (hsSampleData)
data (hsMirrorLocs)

spm1mb <- calcSpm(hsSampleData, hsMirrorLocs)
spm4mb <- calcSpm(hsSampleData, hsMirrorLocs, sigma=4000000)

siglevel1mb <- findSigLevelTrad(hsSampleData, spm1mb, n=3)
siglevel4mb <- findSigLevelTrad(hsSampleData, spm4mb, n=3)

plotScaleSpace(list(spm1mb, spm4mb), list(siglevel1mb, siglevel4mb), type='g')
```

probeAnnotation-class

Class "probeAnnotation"

Description

Holds the probe annotation

Objects from the Class

Instances of this class are not meant to be created by the user

Slots**chromosome:** Chromosome on which the probe is located**maploc:** Location of the probe on the chromosome**name:** Probe name**Methods****[signature(x = "probeAnnotation"):** Access information about a probe**initialize signature(.Object = "probeAnnotation"):** Internal use only**Author(s)**

Jorma de Ronde

```
samplePointMatrix-class
```

Sample point matrix

Description

A sample point matrix resulting from a call to calcSpm

Objects from the Class

Objects can not be created by the user directly but rather through calcSpm.

Slots

totalLength: The total length of the sample point matrix, measures in sample points

maxy: The maximum KC score attained over the sample point matrix

miny: The minimum KC score attained over the sample point matrix

sampleDensity: The sample density used to calculate the sample point matrix. ie the distance between two points in the sample point matrix, measured in base pairs.

sigma: The sigma used for the kernel to calculate the sample point matrix.

mirrorLocs: The mirror locations list used to calculate the sample point matrix

probeAnnotation: The original probe annotation from the input data.

data: The sample point matrix data points in the form of a list where each list element represents a chromosome.

Methods

plot signature(x = "samplePointMatrix"): ...

show signature(object = "samplePointMatrix"): ...

Examples

```
showClass("samplePointMatrix")
```

```
sigSegments-class
```

Significant segments

Description

Lists the significant segments found in a given sample point matrix using a given significance level

Objects from the Class

Objects can not be created by the user directly but rather through getSigSegments.

Slots**gains:** Gained segments**losses:** Lost segments**sigma:** The sigma used for the kernel to calculate the sample point matrix.**sigLevels:** The significance levels at which significant segments are calculated**Methods****show** signature(object = "sigSegments"): ...**write.table** signature(x = "sigSegments"): ...**Examples**

```
showClass("sigSegments")
```

```
write.table           Write summary of the significant regions to a table
```

Description

Write summary of the significant regions to a table

Usage

```
write.table(x, file, append, quote, sep, eol, na, dec, row.names, col.names, qm
## S4 method for signature 'sigSegments':
write.table(x, file="", append = FALSE, quote = 7, sep = "\t", eol = "\n", na =
```

Arguments

x	The sigSegments object to be summarized
file	either a character string naming a file or a connection open for writing. "" indicates output to the console.
append	logical. Only relevant if 'file' is a character string. If 'TRUE', the output is appended to the file. If 'FALSE', any existing file of the name is destroyed.
quote	a logical value ('TRUE' or 'FALSE') or a numeric vector. If 'TRUE', any character or factor columns will be surrounded by double quotes. If a numeric vector, its elements are taken as the indices of columns to quote. In both cases, row and column names are quoted if they are written. If 'FALSE', nothing is quoted.
sep	the field separator string. Values within each row of 'x' are separated by this string.
eol	the character(s) to print at the end of each line (row).
na	the string to use for missing values in the data.
dec	the string to use for decimal points in numeric or complex columns: must be a single character.
row.names	either a logical value indicating whether the row names of 'x' are to be written along with 'x', or a character vector of row names to be written.

<code>col.names</code>	either a logical value indicating whether the column names of 'x' are to be written along with 'x', or a character vector of column names to be written. See the section on 'CSV files' for the meaning of 'col.names = NA'.
<code>qmethod</code>	a character string specifying how to deal with embedded double quote characters when quoting strings. Must be one of "escape" (default), in which case the quote character is escaped in C style by a backslash, or "double", in which case it is doubled. You can specify just the initial letter.

Details

Writes a summary of the sigSegments object to file. The resulting table contains 7 columns. The interpretation of the columns is as follows:

Status Either 'L' for loss or 'G' for gain

Chromosome The chromosome on which this segment is located

Start The start position (in base pairs) of the segment on the chromosome

End The end position of the segment on the chromosome

Average KC score The average KCsmart score over all base pairs in this segment

Mode KC score The highest (for gains) or lowest (for losses) KCsmart score over all base pairs in this segment

Probes All probes from the original data that fall into this segment

Author(s)

Jorma de Ronde

See Also

[calcSpm](#), [getSigSegments](#)

Examples

```
data(hsSampleData)
data(hsMirrorLocs)

spm1mb <- calcSpm(hsSampleData, hsMirrorLocs)

siglevel1mb <- findSigLevelTrad(hsSampleData, spm1mb, n=3)

sigSegments1mb <- getSigSegments(spm1mb, siglevel1mb)
write.table(sigSegments1mb, file=file.path(tempdir(), 'sigSegments1mb.txt'))
```

Index

*Topic **IO**

write.table, 17

*Topic **classes**

KCData-class, 1

KcghData-class, 3

KcghDataMirror-class, 3

KcghDataSplit-class, 4

KcghDataSum-class, 5

probeAnnotation-class, 15

samplePointMatrix-class, 16

sigSegments-class, 16

*Topic **datasets**

hsMirrorLocs, 10

hsSampleData, 11

mmMirrorLocs, 12

*Topic **file**

write.table, 17

*Topic **hplot**

plot, 13

plotScaleSpace, 14

*Topic **iplot**

idPoints, 11

*Topic **manip**

calcSpm, 5

findSigLevelFdr, 7

findSigLevelTrad, 8

getSigSegments, 9

*Topic **package**

KCsmart-package, 2

[, probeAnnotation-method

(probeAnnotation-class), 15

[[, KCData-method (KCData-class), 1

[[<- , KCData-method

(KCData-class), 1

calcSpm, 2, 5, 14, 18

findSigLevelFdr, 2, 7

findSigLevelTrad, 2, 6, 8, 10

getSigSegments, 2, 9, 18

hsMirrorLocs, 10

hsSampleData, 11

idPoints, 11, 14

initialize, KcghData-method

(KcghData-class), 3

initialize, KcghDataMirror-method

(KcghDataMirror-class), 3

initialize, KcghDataSplit-method

(KcghDataSplit-class), 4

initialize, KcghDataSum-method

(KcghDataSum-class), 5

initialize, probeAnnotation-method

(probeAnnotation-class), 15

KCData-class, 1

KcghData-class, 3

KcghDataMirror-class, 3

KcghDataSplit-class, 4

KcghDataSum, 3

KcghDataSum-class, 5

KCsmart (KCsmart-package), 2

KCsmart-package, 2

length, KCData-method

(KCData-class), 1

mmMirrorLocs, 12

plot, 2, 6, 12, 13, 15

plot, samplePointMatrix, missing-method

(plot), 13

plot, scaleSpace, missing-method

(plot), 13

plotScaleSpace, 2, 6, 7, 9, 14, 14

probeAnnotation-class, 15

samplePointMatrix-class, 16

show, samplePointMatrix-method

(samplePointMatrix-class),

16

show, sigSegments-method

(sigSegments-class), 16

sigSegments-class, 16

sort, KcghData, missing-method

(KcghData-class), 3

sort, KcghDataSum, missing-method

(KcghDataSum-class), 5

`unlist`, `KCData`-method
 (*KCData-class*), [1](#)

`write.table`, [10](#), [17](#)

`write.table`, `sigSegments`-method
 (*write.table*), [17](#)