Package 'HiTC'

September 24, 2012

Type Package

Title High Throughput Chromosome Conformation Capture analysis

Description
The HiTC package was developed to explore high-throughput 'C' data such as 5C or Hi-C.

Version 1.0.0

Date 2012-10-01

Depends R (>= 2.10.0), methods, girafe (>= 1.3.1), genomeIntervals (>= 1.7.1), RColorBrewer

Imports methods, Biobase, Biostrings, graphics, grDevices, ShortRead

Suggests

Author Nicolas Servant

Maintainer N. Servant <nicolas.servant@curie.fr>

License Artistic-2.0

Collate AllGenerics.R HTCexp.R qualityControl.R mapC.R normalize.R
binningC.R import.R export.R

LazyLoad yes

biocViews Sequencing, HighThroughputSequencing

R topics documented:

`binningC` *Windowing of high-throughput 'C' interaction matrix*

Description

Windowing of 'C' interaction map

Usage

```
binningC(x, binsize=100000, bin.adjust=TRUE, upa=TRUE,
method="median", use.zero=TRUE, step=1, bnorm=FALSE)
```

Arguments

`x`	object that inherits from class `HTCexp`
`binsize`	size of the bin to consider for windowing
`bin.adjust`	logical; adjust the size of the bin to the size of the genomic region
`upa`	logical; unique primer assignment. Allow one primer to belong to one or several bins
`method`	the method used to combine the counts. Must be 'mean', 'median' or 'sum'
`use.zero`	logical; use the zero values in the `method` calculation
`step`	numeric; binning step size in `n` coverage *i.e.* window step
`bnorm`	logical; normalise each combined counts by the number of primers in the bin

Details

`bin.adjust` allows to work with bin of the same size. Otherwise, the last bin will has a size different from `binsize`. A primer is assigned to a bin, if there is at least one base overlap between the bin and the primer region.
The `method` used to combine the counts in a bin, must be 'mea', 'median' or 'sum'. The `step` parameter allows to choose the overlap between the bins. A `step` of 2 means a 50% overlap between two bins, a step of 3 means a 60% overlap between two bins, *etc.*

Value

An `HTCexp-class` object with binned intraction data. In this case, the primers are converted into bins, and the reverse or forward intervals are similar. The interaction matrix is symetric.

Author(s)

N. Servant, B. Lajoie

See Also

`HTCexp-class`

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Data binning 50kb, with a 4/5 overlap
GM12878.bin <- binningC(GM12878$chr16chr16, binsize=50000, step=5)
show(GM12878.bin)
```

CQC *Quality Control for high-throughput 'C' experiment*

Description

Quality Control for high-throughput 'C' experiment

Usage

```
CQC(x,  cis.trans.ratio = TRUE, hist.interac=TRUE, scat.interac.dist=TRUE,
hist.dist=TRUE, dev.new=FALSE)
```

Arguments

x	object or list of objects that inherits from class HTCexp
cis.trans.ratio	logical; barplot of percentage of inter-intrachromsomal interactions
hist.interac	logical; histogram of the interaction frequency
scat.interac.dist	logical; scatter plot of interaction count versus the genomic distance between two elements
hist.dist	logical; histogram of the distance between the 'x' and 'y' intervals
dev.new	logical; specifying if each plots must be in a separate graphical device

Value

Return NULL; Create quality plots and print some additional informations

Author(s)

N. Servant, B. Lajoie

See Also

HTCexp-class

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Quality Control
CQC(GM12878)
```

discretize *Transform matrix of counts data into discrete matrix*

Description

Transform matrix of counts data into discrete matrix

Usage

```
discretize(x, nb.lev=4, quant=TRUE)
```

Arguments

x	data matrix
nb.lev	number of discretization level
quant	logical; use quantile distribution or split data into equals 'nb.lev' levels

Value

A discrete matrix

Author(s)

N. Servant

See Also

quantile

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Data binning
GM12878bin<-binningC(GM12878$chr16chr16)

## Discretize matrix
dismat<-discretize(intdata(GM12878bin))
mapC(dismat)
```

export.my5C *Export* HTCexp *object to my5C website format*

Description

Export HTCexp object to my5C website format

Usage

```
export.my5C(x, outputfile)
```

Arguments

x	object that inherits from class HTCexp
outputfile	character; the name of the output file

Value

A my5C tabbed delimited file (BED format), with :
Y_INTERVAL_NAME/X_INTERVAL_NAME/INTERACTION_COUNT

Author(s)

N. Servant

See Also

export

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Data binning
GM12878.bin<-binningC(GM12878$chr16chr16, binsize=50000, step=5)

## Export the new intervals definition
export.my5C(GM12878.bin, outputfile="GM12878my5C.csv")
```

`extractRegion` *Extract a subset of the HTCexp object*

Description

Extract a subset of the HTCexp object based on genomic ranges

Usage

```
extractRegion(x, from, to, exact=FALSE)
```

Arguments

x	object that inherits from class HTCexp
from	numeric; start of the genomic region
to	numeric; end of the genomic region
exact	logical; exact genomic region

Details

By default, only the intervals fully included in the genomic ranges are returned. If exact is true, the overlapping intervals are also used, and forced to start/end at the specified position. If no intervals are overlapping, an interval with NA values is added.

Value

A HTCexp object

Author(s)

N. Servant

See Also

Genome_intervals-class, fracOverlap

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Focus on the genomic region chrX:98000000-100000000
GM12878sub<-extractRegion(GM12878$chr16chr16, from=100000, to=300000)
GM12878sub
```

`getExpectedCounts` *Estimate expected interaction counts of a High-Throughput C experiment based on the genomic distance between two loci*

Description

The expected interaction is defined as the linear relationship between the interaction counts and the distance between two primers. See details for addtional informations.

Usage

```
getExpectedCounts(x, span=0.01, bin=0.005, stdev=FALSE, plot=FALSE)
```

Arguments

`x`	object that inherits from class `HTCexp`
`span`	fraction of the data used for smoothing at each x point.
`bin`	interpolation parameter
`stdev`	logical, calculate the variance
`plot`	logical, display loess smoothing and variance estimation points

Details

The estimation of the background is based on the linear interpolation of the counts with the primers distances. A lowess smoothing is used to estimate this linear relationship. Lowess uses robust locally linear fits. A window is placed about each x value; points that are inside the window are weighted so that nearby points get the most weight (tricube weight function). The lowess smoothing has two parameters : span (alpha) and bin (beta). The span corresponds to the fraction of the data used to for smoothing at each x point, i.e. to define the neighboring used for the local smoothing. The bin is the interpolotion parameter, and define the interval size in units corresponding to x. If lowess estimates at two x values within delta of one another, it fits any points between them by linear interpolation. The default is 1% of the range of x. If delta=0 all but identical x values are estimated independently. The bin is used to speed up computation: instead of computing the local polynomial fit at each data point it is not computed for points within delta of the last computed point, and linear interpolation is used to fill in the fitted values for the skipped points. This function may be slow for large numbers of points. Increasing bin should speed things up, as will decreasing span.

The variance is then estimated using the same span and bin parameter, at each interpolation points.

Value

A list with the expected interaction map and the estimated variance

Author(s)

N. Servant, B. Lajoie

See Also

`HTCexp-class`,`normPerZscore`, `normPerExpected`, `lowess`

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Estimate expected interaction from distance between intervals
GM12878.exp<-getExpectedCounts(GM12878$chr16chr16, stdev=TRUE, plot=FALSE)
mapC(GM12878.exp$exp.interaction)
```

HTCexp-class *Class 'HTCexp'*

Description

A class for representing high throughput Chromosome Conformation Capture data from next-generation sequencing experiments.

Objects from the Class

Objects can be created either by:

1. calls of the form `new("HTCexp", intdata, Genome_intervals, Genome_intervals)`.
2. using the auxiliary function `HTCexp` and supplying interaction matrix with x and y intervals definition.

Slots

`intdata`**:** Integer matrix, holding the interaction level between each pairs of 'x-y' intervals. The 'y' intervals must be in rows, and the 'x' in columns.

`ygi`**:** Genomic interval of y intervals; see class `genome_intervals` for details

`xgi`**:** Genomic interval of x intervals; see class `genome_intervals` for details

Methods

detail `signature("HTCexp")`: a more detailed output of the experiment than provided by `show`.

divide comparison of two `signature("HTCexp")` objects. Perform the division of the two interaction matrices on the common 'x' and 'y' intervals. The operation is done only on the common intervals of both objects. If one of the two objects has a count to zero, the divided value will be NA.

export create a BED file with the 'x' and 'y' intervals information.

isBinned return TRUE if the data are binned. The method tests if the 'x' and 'y' genome intervals are the same, if each bin has the same size and if the full genomic range is covered.

isIntraChrom return TRUE if the current `signature("HTCexp")` object contains intrachromosomal interaction data

normPerReads normalize the interaction matrix by the total number of reads of the matrix.

normPerExpected normalize the interaction matrix by the expected number of reads based on the distance between two loci.

normPerZscore normalize the interaction matrix by the zscore calculation, which take into account the expected number of counts and the variance.

plot visualization method; Display an heatmap of the interaction data. Refer to the documentation of `mapC` for more details of the plotting function.

range return the genomic range of the `signature("HTCexp")` object

show summarized output of the experiment, with informations about the data dimension and the genomic region studied.

substract comparison of two `signature("HTCexp")` objects. Perform the substraction of the two interaction matrices on the common 'x' and 'y' intervals. The operation is done only on the common intervals of both objects. If one of the two objects has a count to zero, the divided value will be NA.

Author(s)

Nicolas Servant

See Also

`Genome_intervals-class`, `AlignedGenomeIntervals-class`,

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b
K562<-import.my5C(file.path(exDir,"nsmb.1936-S6.txt"),
                 xgi.bed=file.path(exDir,"Bau_K562_REV.bed"), ygi.bed=file.path(exDir,"Bau_K562_FOR.bed"))

## HTCexp description
show(GM12878)
detail(GM12878)

## Is binned data ?
isBinned(GM12878$chr16chr16)

## Is a inter or intrachromsomal experiment ?
isIntraChrom(GM12878$chr16chr16)

## Plotting
plot(GM12878$chr16chr16)
plot(GM12878$chr16chr16, view=2)
plot(binningC(GM12878$chr16chr16), binningC(K562$chr16chr16), maxrange=20)

## Zscore Normalization
GM12878norm<-normPerZscore(GM12878$chr16chr16)

## Operation on HTCexp object
GM12878_d_K562<-divide(normPerReads(GM12878$chr16chr16), normPerReads(K562$chr16chr16))
GM12878_s_K562<-substract(normPerReads(GM12878$chr16chr16), normPerReads(K562$chr16chr16))

## Overlap with genomic annotation
Refgene <- readBED(file.path(system.file("extdata", package="HiTC"),"refseq_hg19_chr16_1_500000.bed"))
plot(GM12878$chr16chr16, giblocs=list(RefSeqGene=Refgene$Refseq_Gene))
```

`import.my5C` *Import data from my5C webtool*

Description

Import data from my5C webtool

Usage

```
import.my5C(my5C.datafile, xgi.bed, ygi.bed, all.pairwise=TRUE)
```

Arguments

`my5C.datafile`	input file from the my5C webtool
`xgi.bed`	BED file describing the 'x' Intervals (i.e. column names) of the interaction map
`ygi.bed`	BED file describing the 'y' intervals (i.e. row names) of the interaction map
`all.pairwise`	logical; generate all pairwise chromosomal interaction maps

Details

The list format from the my5C webtool is a tabbed delimited format (BED format), with : FORWARD_PRIMER_NAME/REVERSE_PRIMER_NAME/INTERACTION_COUNT

The matrix format is tabbed delimited format, corresponding to the interaction map. The rownames and columnames are splitted using the "|" separator (example : REV_2|mm9|chrX:98831149-98834145). The rownames and colnames are then intersected with the ids of the intervals defined in the BED files.

The BED format is a standard format provided by the the UCSC Genome Browser.

The `all.pairwise` option is not necessary in case of symetric design. Otherwise, it will return all the pairwise interaction maps.

Value

A list of `HTCexp` object(s)

Author(s)

N. Servant

See Also

`Genome_intervals-class`, `HTCexp-class`

Examples

```
exDir <- system.file("extdata", package="HiTC")
hiC<-import.my5C(file.path(exDir,"HIC_gm06690_chr14_chr14_1000000_obs.txt"), xgi.bed=file.path(exDir,"GSE18
hiC
```

intervalsDist *intervalsDist*

Description

Compute the distance between the intervals of a 'C' experiment

Usage

```
intervalsDist(x)
```

Arguments

x	object that inherits from class HTCexp

Details

If A and B are the two sets of primers and s and e, the start and end of a primer, the distance is calculated as :

$$\min(|A_e - B_s|, |A_s - B_e|)$$

Value

A matrix of distances between primers

Author(s)

N. Servant

See Also

HTCexp-class

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Calculate distances between primers/intervals
intervalsDist(GM12878$chr16chr16)
```

mapC *Visualize 'C' interaction map*

Description

Visualize 'C' interaction counts matrix

Usage

```
mapC(x, y=NULL, view=1, giblocs=NULL, minrange=NA, maxrange=NA, trim.range=0.98, names=FALSE, va
```

Arguments

x	object that inherits from class HTCexp or from class matrix
y	optional. object that inherits from class HTCexp or from class matrix. If specified, view is set to 2
view	interaction map representation. See details
giblocs	genomeIntervals object of blocks to display as annotation track(s)
minrange	the minimum range of values used to define the color palette
maxrange	the maximum range of values used to define the color palette
trim.range	define the maxrange and minrange values using the percentile of the interaction matrix.
names	logical; display the names of the intervals. Useful for small matrices
value	logical; display the interaction values on the matrix. Useful for small matrices
show.na	logical; show the NA values in gray
log.data	logical; do you want to log the data before plotting the heatmap
col.pos	color for (low,mid,high) positive interaction counts. Must be a vectore of size 3. mid can be NA
col.neg	color for (low,mid,high) negative interaction counts. Must be a vectore of size 3. mid can be NA
col.na	color for NA values
mask.data	matrix to add to the heatmap as a mask. Must have the same dimension as the interaction matrix
grid	logical; add a grid on the heatmap
title	character; add a title to the heatmap

Details

This function implements the plot method for objects of class HTCexp.

By default, the maxrange and minrange values are fixed as the 98th percentile (resp. 2th percentile) of the interaction matrix. These values are useful to play with the contrast and remove the extreme values from the matrix.

Two different views are available. The heatmap view (view=1) display the data in two dimension. The triangle view (view=2) only represent the top-right part the interaction matrix. If two HTCexp objects are specified the view is force to 2, in order to compare both interaction maps. The two maps have to be binned to ensure comparison between genomic ranges.

Annotation tracks can be added to both views. In case of binned data, the exact genomic positions of each features are takken into account. Otherwise, the 'C' intervals which overlap with the annotation features are colored.

Value

Returns NULL; this function is called for the side-effect of creating the plot.

Author(s)

N. Servant, B. Lajoie

See Also

interval_overlap

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b

## Interaction map
mapC(GM12878$chr16chr16)

## Play with contrast and color
mapC(GM12878$chr16chr16, maxrange=100, col.pos=c("black","red","yellow"))

## Add annotation and change view
exDir <- system.file("extdata", package="HiTC")
gene <- readBED(file.path(exDir,"refseq_hg19_chr16_1_500000.bed"))
mapC(GM12878$chr16chr16, giblocs=list(Refseq=gene$Refseq_Gene), view=2)

## Compare two samples
K562<-import.my5C(file.path(exDir,"nsmb.1936-S6.txt"),
                 xgi.bed=file.path(exDir,"Bau_K562_REV.bed"), ygi.bed=file.path(exDir,"Bau_K562_FOR.bed"))
mapC(binningC(GM12878$chr16chr16, binsize=50000, step=5),
binningC(K562$chr16chr16, binsize=50000, step=5), giblocs=list(Refseq=gene$Refseq_Gene))
```

---

readBED *readBED*

Description

read BED files and convert tracks in genomeIntervals objects

Usage

```
readBED(con)
```

Arguments

con	BED file to read

Details

If a score column is specified in the BED file, it will be saved as a 'score' annotation slot.

Value

A list of GenomeIntervals object(s). Each element corresponds to a track.

Author(s)

N. Servant

See Also

Genome_intervals-class

Examples

```
exDir <- system.file("extdata", package="HiTC")
gene <- readBED(file.path(exDir,"refseq_hg19_chr16_1_500000.bed"))
```

---

removeIntervals *Remove intervals from HTC object*

Description

Remove primers intervals from HTC object

Usage

```
removeIntervals(x, ids)
```

Arguments

x	object that inherits from class HTCexp
ids	character; vector of primers Ids to remove from the object

Value

A HTCexp object without the discarded intervals

Author(s)

N. Servant

See Also

Genome_intervals-class

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b


## Remove intervals from a HTCexp object
removeIntervals(GM12878$chr16chr16, ids=c("5C_305_ENm008_FOR_7", "5C_305_ENm008_REV_60"))
```

---

setIntervalScale *Set x and y interval of the HTCexp object*

Description

Set x and y interval of the HTCexp object and update the interaction map accordingly

Usage

```
setIntervalScale(x, xgi, ygi, upa=TRUE, method="mean", use.zero=TRUE)
```

Arguments

x	object that inherits from class HTCexp
ygi	y intervals; see class genome_intervals for details
xgi	x intervals; see class genome_intervals for details
upa	logical; unique primer assignment. Allow one primer to belong to one or several bins
method	the method used to combine the counts. Must be 'mean', 'median' or 'sum'
use.zero	logical; use the zero values in the method calculation

Details

Define new interaction map based on the specified xgi and ygi intervals.

This function has to be used carefully and can has important impact on the interaction map. It is important to note that the setIntervalScale function is different from the binningC function in the way that the output is not symetrical.

Value

A HTCexp object

Author(s)

N. Servant

See Also

HTCexp-class

Examples

```
exDir <- system.file("extdata", package="HiTC")
GM12878<-import.my5C(file.path(exDir,"nsmb.1936-S5.txt"),
                 xgi.bed=file.path(exDir,"Bau_GM12878_REV.bed"), ygi.bed=file.path(exDir,"Bau_GM12878_FOR.b
GM12878.bin<-binningC(GM12878$chr16chr16)

K562<-import.my5C(file.path(exDir,"nsmb.1936-S6.txt"),
                 xgi.bed=file.path(exDir,"Bau_K562_REV.bed"), ygi.bed=file.path(exDir,"Bau_K562_FOR.bed"))

## I have two HTCexp samples defined with different intervals.
show(GM12878.bin)
show(K562$chr16chr16)

## How to compare them ?
## One idea is to force the intervals definition of one object using the
## intervals of the other.

setIntervalScale(K562$chr16chr16, xgi=x_intervals(GM12878.bin), ygi=y_intervals(GM12878.bin))
```

Index