Package 'HiTC'

March 26, 2013

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 1.2.0

Date 2012-09-21

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 Nicolas 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 deprecated.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)
```

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

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 is 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

```
data(Nora_5C)

## Data binning 100kb, with a 1/3 overlap
E14.bin <- binningC(E14$chrXchrX, binsize=100000, step=3)
show(E14.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, trim.range=0.98, 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
trim.range	remove the extreme values by trimming the counts. Only use for plotting functions. [0,1]
dev.new	logical; specifying if each plots must be in a separate graphical device

Details

If x is a list, all HTCexp objects are merged. The zero values are not used to compute the descriptive statistics and to display the data. If trim.range are lower than 1. The highest values (quantile probability is equal to trim.range) are discarded.

Value

Return a matrix; Create quality plots and return a matrix with some simple statistics on all, cis and trans data.

Author(s)

N. Servant, B. Lajoie

See Also

HTCexp-class

Examples

```
data(Nora_5C)

## Quality Control
CQC(E14)
```

---

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

```
data(Nora_5C)

## Data binning
E14bin<-binningC(E14$chrXchrX)

## Discretize matrix
dismat<-discretize(intdata(E14bin))
mapC(dismat)
```

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

Description

Export HTCexp object to my5C website format

Usage

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

Arguments

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

Value

A my5C tabbed delimited file, with :
Y_INTERVAL_NAME/X_INTERVAL_NAME/INTERACTION_COUNT

Author(s)

N. Servant

See Also

exportC

Examples

```
data(Nora_5C)

## Data binning
E14.bin<-binningC(E14$chrXchrX)

## Export the new intervals definition
export.my5C(E14.bin, file="E14my5C.csv")
```

exportC *Export* HTCexp *object*

Description

Export HTCexp object to csv format

Usage

```
exportC(x, file)
```

Arguments

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

Value

A csv file, with :
chrA,startA,endA,nameA,strandA,chrB,startB,endB,nameB,strandB,countAB

Author(s)

N. Servant

See Also

export.my5C, importC

Examples

```
data(Nora_5C)

## Data binning
E14.bin<-binningC(E14$chrXchrX)

## Export the new intervals definition
exportC(E14.bin, file="E14.csv")
```

---

extractRegion *Extract a subset of the HTCexp object*

Description

Extract a subset of the HTCexp object based on genomic ranges

Usage

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

Arguments

x	object that inherits from class HTCexp
chr	character; the chromosome of the genomic region
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

```
data(Nora_5C)

## Focus on the genomic region chrX:98000000-100000000
E14sub<-extractRegion(E14$chrXchrX, chr="chrX", from=98000000, to=100000000)
E14sub
```

---

getExpectedCounts *Estimate expected interaction counts of a High-Throughput C intra-chromsomal map 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 loci. See details for additional 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 lowess smoothing and variance estimation points

Details

The expected value is the interaction frequency between two loci that one would expect based on a sole dependency on the genomic proximity of these fragments in the linear genome. This can be estimated using a Lowess regression model. The lowess smoothing has two parameters : span and bin. The span corresponds to the fraction of the data used for smoothing. Instead of computing the local polynomial fitting at each data point, a window of size delta (bin parameter) is applied on the data and a linear interpolation is used to fill in the fitted values within the window. The default is 1% of the range of x. If delta=0 all but identical x values are estimated independently. The variance is then estimated using the same span and bin parameter, at each interpolation points. The points inside a window are weighted so that nearby points get the most weight (tricube weight function).

Value

A list with the expected interaction map and the estimated variance

Author(s)

N. Servant, B. Lajoie

See Also

HTCexp-class,normPerExpected, normPerExpected, lowess

Examples

```
data(Nora_5C)

## Estimate expected interaction from distance between intervals
E14.exp<-getExpectedCounts(E14$chrXchrX, stdev=TRUE, plot=FALSE)
mapC(E14.exp$exp.interaction)
```

HTCexp-class *Class 'HTCexp'*

Description

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

Details

The normPerExpected method estimates the expected interactions based on a the dependency on the genomic proximity between two loci. Look at the getExpectedCounts function for details.

The normPerTrans method is based on the assumption that all trans interactions should be the same. Thus, the cis interactions can be normalized by the interaction level of trans data. The xtrans trans map has to share its 'xgi' intervals with the cis map, and the ytrans has to share its 'ygi' intervals with the cismap. The method is used to combine the normalization factor from x and y intervals. Must be 'sum', 'mult' or 'mean'.

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 Deprecated. See exportC function

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 Depracted. See normPerExpected

normPerTrans Normalize cis interaction map based on the trans interactions. see details.

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

```
data(Nora_5C)

## HTCexp descriptio
show(E14)
detail(E14)

## Is binned data ?
isBinned(E14$chrXchrX)

## Is a inter or intrachromsomal experiment ?
isIntraChrom(E14$chrXchrX)

## Plotting
plot(E14$chrXchrX)
```

```
plot(E14$chrXchrX, view=2)
plot(binningC(E14$chrXchrX), binningC(MEF$chrXchrX), maxrange=20)

## Divide by expected interaction counts
E14norm<-normPerExpected(E14$chrXchrX)

## Operation on HTCexp object
E14_d_MEF<-divide(normPerReads(E14$chrXchrX), normPerReads(MEF$chrXchrX))
E14_s_MEF<-substract(normPerReads(E14$chrXchrX), normPerReads(MEF$chrXchrX))

## Overlap with genomic annotation
Refgene <- readBED(file.path(system.file("extdata", package="HiTC"),"refseq_mm9_chrX_98831149_103425150.bed
plot(E14$chrXchrX, giblocs=list(RefSeqGene=Refgene$Refseq_Gene))

## Not run:
## normPerTrans data normalization applied on \href{http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEnco
ENCODE=import.my5C("./ENM-GM12878-R1.matrix")

## Look at raw interaction map
mapC(ENCODE$chr7chr7)

## look at normalize by trans interaction map
mapC(normPerTrans(ENCODE$chr7chr7, xtrans=ENCODE$chr7chr5, ytrans=ENCODE$chr5chr7))

## End(Not run)

## Not run:
## Export
exportC(E14$chrXchrX, con="E14.csv")

## End(Not run)
```

---

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. Required for the my5C list format
ygi.bed	BED file describing the 'y' intervals (i.e. row names) of the interaction map. Required for the my5C list format
all.pairwise	logical; generate all pairwise chromosomal interaction maps, i.e chr1-chr2, chr2-chr1

Details

This function allows data import from the the my5C webtool. Two input formats can be used : - The list format is composed of three files; two BED files describing the genomic intervals (i.e. primers); and a tabbed delimited format to specify the interaction between each genomic regions, with : FORWARD_PRIMER_NAME/REVERSE_PRIMER_NAME/INTERACTION_COUNT

- The matrix format is a tab-delimited format, corresponding to the interaction map. The rownames and columnames are splitted to created the genome intervals (example : REV_2|mm9|chrX:98831149-98834145).

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")
## Load my5C matrix format
hiC<-import.my5C(file.path(exDir,"HIC_gm06690_chr14_chr14_1000000_obs.txt"))
hiC
```

---

importC *Import high-htroughput 'C' data*

Description

Import 5C or Hi-C data from csv file

Usage

```
importC(con, all.pairwise=TRUE)
```

Arguments

con	input csv file. See details
all.pairwise	logical; generate all pairwise chromosomal interaction maps, i.e chr1-chr2, chr2-chr1

Details

This function import high-throughput data from a csv file. The expected format is the following : chrA,startA,endA,nameA,strandA,chrB,startB,endB,nameB,strandB,countAB

Value

A list of HTCexp object(s)

Author(s)

N. Servant

See Also

exportC,import.my5C, HTCexp-class

Examples

```
data(Nora_5C)

## Data binning
E14.bin<-binningC(E14$chrXchrX)

## Export the new intervals definition
exportC(E14.bin, file="E14.csv")

##Import
importC("E14.csv")
```

---

intervalsDist *intervalsDist*

Description

Compute the distance of intrachromosomal interactions 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 :

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

Only intrachromsomal interaction maps can be use for this operation.

Value

A matrix of distances between primers

Author(s)

N. Servant

See Also

HTCexp-class

Examples

```
data(Nora_5C)

## Calculate distances between primers/intervals
intervalsDist(E14$chrXchrX)
```

---

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,
```

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

```
data(Nora_5C)

## Interaction map
mapC(E14$chrXchrX)

## Play with contrast and color
mapC(E14$chrXchrX, 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_mm9_chrX_98831149_103425150.bed"))
mapC(E14$chrXchrX, giblocs=list(Refseq=gene$Refseq_Gene), view=2)

## Compare two samples
mapC(binningC(E14$chrXchrX), binningC(MEF$chrXchrX), giblocs=list(Refseq=gene$Refseq_Gene))
```

Nora_5C *HiTC - 5C data*

Description

5C data described by Nora et al. (2012)

Usage

```
data(Nora_5C)
```

Format

Contains two list of HTCexp objects (E14 and MEF). Data from the chromosome X are available.

Details

This 5C dataset published by Nora et al (GSE35721), contains two different samples, a male undifferentiated ES cells (E14, GSM873935) and a mouse embryonic fibroblasts (MEF, GSM873924). This dataset is mainly used to describe the available functionalities of the HiTC package. The data provided with the package are count data.

Source

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE35721

References

Nora EP, Lajoie BR, Schulz EG, Giorgetti L et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature 2012 Apr 11;485(7398):381-5. PMID: 22495304

Examples

```
data(Nora_5C)
show(E14)
show(MEF)
```

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_mm9_chrX_98831149_103425150.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

```
data(Nora_5C)

## Remove intervals from a HTCexp object
removeIntervals(E14$chrXchrX, ids=c("5C_938_XIC-3_REV_2", "5C_938_XIC-3_REV_4"))
```

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

```
data(Nora_5C)

E14.bin<-binningC(E14$chrXchrX)

## I have two HTCexp samples defined with different intervals.
show(E14.bin)
show(MEF$chrXchrX)

## How to compare them ?
```

```
## One idea is to force the intervals definition of one object using the
## intervals of the other.

setIntervalScale(MEF$chrXchrX, xgi=x_intervals(E14.bin), ygi=y_intervals(E14.bin))
```

Index