Package 'R3CPET'

October 9, 2015

Type Package
Title 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process
Description The package provides a method to infer the set of proteins that are more probably to work together to maintain chormatin interaction given a ChIA-PET experiment results.
Version 1.0.0
Date 2013-11-23
Author M. N. Djekidel, Yang Chen et al.
Maintainer Mohamed Nadhir Djekidel <djek.nad@gmail.com></djek.nad@gmail.com>
License GPL (>=2)
Depends R ($>= 3.2$), Rcpp ($>= 0.10.4$), methods
Imports parallel, clues, ggplot2, pheatmap, IRanges, clValid, igraph,data.table,GenomicRanges,S4Vectors, DAVIDQuery, ggbio,reshape2,Hmisc
LinkingTo Rcpp
VignetteBuilder knitr
Suggests BiocStyle, knitr, TxDb.Hsapiens.UCSC.hg19.knownGene, biovizBase, biomaRt, AnnotationDbi, org.Hs.eg.db, shiny, ChIPpeakAnno
biocViews NetworkInference, GenePrediction, Bayesian, GraphAndNetwork, Network, GeneExpression
NeedsCompilation yes
R topics documented:
R3CPET-package annotateExpression-methods Biogrid buildNetworks ChiapetExperimentData-class

2 R3CPET-package

visualizeInteractions	15
	14
F	13
RPKMS	12
PrepareData-methods	11
	10
	38
outputGenesPerNetworkToDir	36
outputGenesPerClusterToDir-methods	35
NetworkCollection-class	34
loadTFBS-methods	33
	31
	29
	27
HPRD	27
	26
	24
	23
	21
	20
	20
	9
	18
	16
	15
	13
	12
	10
chromosoms	1

Description

The main goal of 3CPET is to try to infer the set of protein networks that are likely to work together inorder to maintain chromatin loops obtained by a ChIA-PET experiment. It is based on an idea silimar to the one used for document classification. It starts first by building a PPI network for each chromatin interaction, then uses an HDLA (Hierarchical Dirichlet Latent Allocation) model to infer the set of networks that are enriched together.

a Hierarchical Dirichlet Process

Details

Package: R3CPET
Type: Package
Version: 1.0
Pate: 2012 11 2

Date: 2013-11-23 License: GPL (>= 3.0)

Author(s)

Written by M.N.Djekidel Maintainer: Mohamed Nadhir Djekidel <nde12@mails.tsinghua.edu.cn>

References

M.N Djekidel et al, 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process, in press, 2015

See Also

ChiapetExperimentData, ChromMaintainers, HLDAResult

annotateExpression-methods

Add the gene expression attribute to the graph nodes

Description

This method is a kinda of helper method, it helps the user to add for each node in the inferred chromatin maintainer network the RPKM attributes. It is useful if the user want to save the networks as ".gml" files and visualize them using software such as Gephi or Cytoscape. Or maybe if he wants to know which networks are highly expressed then others.

Usage

```
## S4 method for signature 'ChromMaintainers,data.frame'
annotateExpression(object, RPKMS)
```

Arguments

object a ChromMaintainers object in which the networks are already present

RPKMS a two columns data. frame, where the first column contains the name of the

gene and the second contains the expression values

4 Biogrid

Value

A ChromMaintainers object in which the networks are annotated.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

ChromMaintainers, InferNetworks

Examples

```
## get the different datasets path
  petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")</pre>
  tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")
## Not run:
 x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
  ## build the diffrent indexes
  x <- createIndexes(x)</pre>
## build the different indexes
x <- createIndexes(x)</pre>
## build networks connecting each interacting regions
nets<- buildNetworks(x)</pre>
## infer the networks
hlda<- InferNetworks(nets)</pre>
networks(hlda)
## Annotate networks
hlda<- annotateExpression(hlda,as.data.frame(RPKMS))</pre>
## Notice the addition of the RPKM attribute to each network
networks(hlda)
## End(Not run)
```

Biogrid

Biogrid Network

Description

loads an igraph object that contains the Biogrid V 2.0.49 PPI.

buildNetworks 5

Usage

```
data(Biogrid)
```

Value

an igraph named PPI.Biogrid.

Source

http://thebiogrid.org/

Examples

```
data(Biogrid)
PPI.Biogrid
```

buildNetworks

Building interaction networks connecting interacting regions

Description

This methods uses the background PPI to try to build an interaction network that connects each interacting regions. If a regionA interacts with a regionB and if TF_A is the list of TF in regionA and TF_B is the list of TF in regionB, than we use the loaded PPI as a background network to connect each TF from TF_A to each TF in TF_B .

We suppose that a minimum number of physical interactions (minimum energy) are needed to connection each TF to the other. Thus, we take the shortest path in the PPI. at this stage, each network is a collection of edges.

Usage

```
## S4 method for signature 'ChiapetExperimentData'
buildNetworks(object, minFreq = 0.25, maxFreq = 0.75)
```

Arguments

object	a ChiapetExperimentData object in which the interactions and TFBS and PPI are already loaded. Check loadPETs, loadTFBS, loadPPI for more info.
minFreq	After constructing the networks for all the interacting regions all edges that appear in less than minFreq of the networks are considered to be outliers.
maxFreq	After constructing the networks for all the interacting regions all edges that appear in more than maxFreq of the networks are considered to be interactions involving general TF and are removed.

Value

A NetworkCollection object that contain the list of all the constructed networks and their sizes.

NOTE: interactions for which no TF was bound or no networks could be constructed or which was empty after filtering will not be considered.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

ChiapetExperimentData, loadTFBS, loadPETs, loadPPI, createIndexes

Examples

```
## get the different datasets path
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:

x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
## build the diffrent indexes
x <- createIndexes(x)

## build networks connecting each interacting regions
nets<- buildNetworks(x)
nets

## End(Not run)</pre>
```

ChiapetExperimentData-class

3CPET used raw data

Description

The ChiapetExperimentData class is a container for storing the set of raw data used by 3CPET to do the prediction.

Usage

Arguments

pet (optional) a ChIAP-PET interactions file path or a GRanges object. The GRanges

object should have a column named PET_ID. details on the file format can be

found on the loadPETs help page.

tfbs (optional) a file path to a BED file containing transcription factors binding site

or a GRanges object. The GRanges object should have a metadata column named

TF.

ppi (optional) an igraph object that contains a user defined protein-protein interac-

tion network. if this parameter is not specified, the ppiType paramter will be

used.

IsBed (optional) considered only if the pet parameter is a file path. More info about

this paramters can be found in the loadPETs help page.

petHasHeader (optional) logical. Indicates if the ChIA-PET interactions file has a header or

not.

dist (optional) The size of the region to consider arround the center of the interacting

regions.

tfbsHasHeader (optional) logical. Indicates if the TFBS file has a header or not.

ppiType (optional) considered only if the ppi paramter is not specified. This paramter tell

the pakage to load one of the PPI (HPRD or Biogrid) shiped with the package.

filter (optional) logical. whether of not to filter the PPI network. if the RPKM

paramter is specified then the RPKM dataset incorporated with the package will be used. if you want to to your own way of filtering, you ca set filter = FALSE

and pass an already processed PPI to the ppi paramterer.

term (optional) the GO term used to filter the nodes of the PPI. this is different from

the filter parameter in the filter parameter the PPI nodes are filtered by their gene expression, while in the term parameter they are filtered by their

genomic location. by default "GO: 0005634" is used for filtering.

annot (optional) the annotation dataset used for filtering by default the geneLocations

is used. The user can pass a custom data.frame. For more details check the

loadPPI help page.

RPKM (optional) a data. frame object that contains the expression value of each genes.

by default the RPKMS dataset will be used (expression value in K562 celline). For

more information about the format of the data passed to this parameter please check the loadPPI

threshold (optional) threshold value used to filter gene expression. default: 1.

Details

The ChiapetExperimentData class stores the genomic coordinates of the ChIA-PET interactions, the binding sites of the different transcription factor (TFBS) and the background protein-protein interaction (PPI) network used to infer the final chromatin maintainer networks.

Value

Constructs a ChiapetExperimentData object with the specified fields populated.

Slots

- **pet**: Object of class GRanges that stores the genomic coordinated of the interactions. it can be populated using the method loadPETs
- **tfbs**: Object of class GRanges that stores the TF binding site. it can be populated using the method loadTFBS. NOTE: the TFBS locations can be obtained from a ChIP-Seq experiment or a motif finding software. for more information on the format of the provided data check loadTFBS
- **ppi** Object of class "igraph" used as the background PPI for further analysis. it can be populated using the method loadPPI
- .dt Object of class "list" contains a collection of data.table serving as indexes used internally by the package (not expected to be manipulated by the user). it can be populated using the method createIndexes

Accessors

The following methods can be used to get the content of a ChiapetExperimentData object x:

pet(x), pet(x) <- value: Get ChIA-PET interactions encoded as a GRanges object in x. The
 returned GRanges objects contains an attribute PET_ID in which the left side have an id of the
 form PET#\d+\.1 and the right side interaction have an id of the form PET#\d+\.2. for more
 information check loadPETs</pre>

	seqnames		ranges	strand		PET_ID
	<rle></rle>		<iranges></iranges>	<rle></rle>	1	<character></character>
[1]	chr1	[1240734,	1242734]	*	1	PET#1.1
[2]	chr1	[1242224,	1244224]	*	1	PET#1.2
[3]	chr1	[1282208,	1284208]	*	1	PET#2.1
[4]	chr1	[1283334,	1285334]	*	1	PET#2.2
[5]	chr1	[1370371,	1372371]	*	1	PET#3.1
[6]	chr1	[1371822,	1373822]	*	I	PET#3.2

tfbs(x), tfbs(x) <- value: Get the GRanges storing the transcription factor binding sites.
ppi(x), ppi(x) <- value: Returns an igraph object used as a background PPI. check the
loadPPI for more information.</pre>

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Li G, Fullwood MJ, Xu H et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biology 2010, 11(2):R22

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

```
loadPETs, loadTFBS, loadPPI
```

```
## for example Reading ChIA-PET interaction results generated from ChIA-PET tool
## it should be formatted as follow:
## chromleft startleft endleft chromright startright endright counts pvalue qvalue
       chr1 872113 879175 chr1 933836 938416 12 1.84529e-30 6.90983e-28
##
      chr1 874165 879175 chr1 933340 938306 10 1.23139e-25 3.58932e-23 chr1 889676 896594 chr1 933897 938982 13 4.91311e-36 2.33753e-33 chr1 898753 907581 chr1 931133 939571 19 0.00000e+00 0.00000e+00 chr1 910103 918775 chr1 930834 938627 15 2.20004e-43 1.32812e-40 chr1 919314 922154 chr1 934212 937864 6 3.70292e-21 7.88551e-19
##
##
##
## The counts, pvalue and qvalue fields are not considered in our case
## it is up to the user to filter the interactions.
## The TFBS should be a BED file that contain the chromosome, start, end and the TF name
## Not run:
  ## load the different datasets
  petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")</pre>
  tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")
 x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
  ## build the diffrent indexes
  x <- createIndexes(x)</pre>
  ## Pass objects instead of files.
  pet <- sample(pet(x), size = 20, replace = TRUE )</pre>
  tfbs <- sample(tfbs(x), size=300, replace=TRUE)
  ppi <- ppi(x)
  tst <- ChiapetExperimentData(pet = pet, tfbs= tfbs, ppi=ppi)</pre>
  tst <- createIndexes(tst)</pre>
```

10 ChromMaintainers-class

```
tst
## End(Not run)
```

ChromMaintainers-class

Chomatin maintainer networks

Description

The ChromMaintainers holds information about the inferred network by the method InferNetworks. It contains the list of inferred networks as igraph object, a list of edges and a list of proteins. In addition to an HLDAResult object that contains the final probabilities calculated by the HLDA algorithm.

Usage

ChromMaintainers(maintainers, topEdges, topNodes, clusRes = NULL, networks = list())

Arguments

maintainers Object of class "HLDAResult" that contains the HLDA results.

topEdges a "matrix" containing the list the top edges per network.

topNodes a "matrix" containing the list the top proteins per network.

networks the list of infered networks as an igraph objects list

clusRes Object of class "cluesOrSota" describing the assignment of each DNA-interaction

to a chromatin-maintainer network according to their enrichment.

Value

 $a \; {\tt ChromMaintainers} \; object.$

Accesors

if x is a ChromMaintainers object the following accessors can be applied:

networks(x) gets the list of networks as igraph objects

topNodes(x) gets a matrix object that contains the list of top proteins per network

topEdges(x) gets a matrix object that contains the list of top proteins per network

getClusters(x) returns the clustering results of DNA-interaction into groups according to their partnership enrichment profile to the set of inferred chromatin maintainer networks.

ChromMaintainers-class 11

Methods

Many plotting and annotation methods are associated with this class.

annotateExpression(object, RPKMS) To add the gene expression attribute to the igraph objects

clusterInteractions(object, method, nbClus) To cluster the DNA-interactions according to their partnership enrichment profile.

GenerateGmlNetworks(object,...) Creates the list of igraph object from the topEdges matrix

outputGenesPerClusterToDir(hdaRes,data,path, ...) get the list genes belonging to each DNA-interaction cluster.

getRegionsIncluster(hdaRes,data, cluster, ...) returns the coordinates of the DNA interactions for a given cluster.

GOEnrich.networks(object, pval=0.05, GOlimit= 5,path="") do a GO enrichment of the elements of each inferred network.

plot3CPETRes(object, path, W, H, type,byEdge, netPerRow, layoutfct, ...) provide different type of plots to visualize the results

visualizeCircos(object, data, cluster, chrLenghts) Draws a circos plot of the DNA interactions in a given cluster.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

```
https://www.cs.princeton.edu/~blei/topicmodeling.html (C. Wang's hdp code)
```

Chong Wang, John Paisley and David M. Blei, *Online variational inference for the hierarchical Dirichlet process*. In AISTATS 2011

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

InferNetworks, ChromMaintainers, HLDAResult

```
showClass("ChromMaintainers")
```

12 cluesOrSota-class

chromosoms

Human chromosom lenghts

Description

This dataset contains the human chromosoms lengths

Examples

data(chromosoms)
Chromosoms

cluesOrSota-class

Wrapper for clues and sota S3 classes

Description

This is an S4 virtual union class that defines a new object that can be a sota or a clues class.

Definition

```
setClassUnion("cluesOrSota", c("clues", "sota","NULL"))
```

Objects from the Class

A virtual Class: No objects may be created from it.

Methods

No methods defined with class "cluesOrSota" in the signature.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Herrero, J., Valencia, A, and Dopazo, J. (2005). A hierarchical unsupervised growing neural network for clustering gene expression patterns. Bioinformatics, 17, 126-136. Wang, S., Qiu, W., and Zamar, R. H. (2007). CLUES: A non-parametric clustering method based on local shrinking. Computational Statistics & Data Analysis, Vol. 52, issue 1, pages 286-298.

See Also

ChromMaintainers

```
showClass("cluesOrSota")
```

clusterInteractions-methods 13

clusterInteractions-methods

Grouping DNA interactions by enrichment profile

Description

This method aims at clustering the DNA interactions according to their partnership probability to the inferred chromatin maintainer networks.

Two kinds of clustering are supported supervised and un-supervised. In the first one the functionsota from the clvalid package. In the second case the clues method from the clues package is used.

Usage

```
## S4 method for signature 'ChromMaintainers'
clusterInteractions(object, method=c("clues","sota"), nbClus=20 )
```

Arguments

Sumones	
object	(Required) a non-empty ChromMaintainers object
method	(<i>optional</i>)used to specify the method to use. if clues is specified, the number of clusters will be determined automatically. if method = "sota", the user can specify the number of clusters by setting the parameter nbClus, by default it is set to 20. by default method = "clues".
	Another option is that user can first do an automatic clustering using "clues" and then if judges that more or less clusters are required he can recall the clusterInteractions method and specify the number the number of clusters.
nbClus	(<i>optional</i>) The user-specified number of clusters. It is taken into consideration only if method = sota.

Value

A ChromMaintainers object in which the clusRes is populated as a sota or clues object.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Herrero, J., Valencia, A, and Dopazo, J. (2005). A hierarchical unsupervised growing neural network for clustering gene expression patterns. Bioinformatics, 17, 126-136.

Wang, S., Qiu, W., and Zamar, R. H. (2007). *CLUES: A non-parametric clustering method based on local shrinking*. Computational Statistics & Data Analysis, Vol. 52, issue 1, pages 286-298

See Also

ChromMaintainers, clues, sota, InferNetworks

Examples

```
data(RPKMS)
 ## get the different datasets path
 petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")</pre>
 tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")</pre>
## Not run:
 x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
 ## build the diffrent indexes
 x <- createIndexes(x)</pre>
 ## build networks connecting each interacting regions
 nets<- buildNetworks(x)</pre>
 ## infer the networks
 hlda<- InferNetworks(nets)</pre>
 #cluster
 hlda<- clusterInteractions(hlda)</pre>
 #Display heatmap
 plot3CPETRes(hlda,type="heatmap")
 hlda
## End(Not run)
```

CreateCenteredBED-methods

Create centered interactions

Description

This helper method can be used to create a "bed" file in which the coordinates of the regions are the centre of the interactions in the raw data. in **R3CPET** we suppose that the centre of the interactions are the most enriched when doing read mapping ,thus, we consider just the region around the centre to detect the TFBS.

Usage

```
## S4 method for signature 'character'
CreateCenteredBED(file, header=TRUE,dist=1000)
```

createIndexes-methods 15

Arguments

file	character indicating the location of the rawdata file, the file should be a s	six

column "bed" file in which the fist 3 columns indicate the left side interaction

(chr, start, stop) and the other 3 columns indicate the right side interaction.

header logical, indicates if the first line in the file is a header.

dist numeric, indicated the distance around the center of the region to take.

Value

A 4 columns data. frame object, in which the first 3 columns indicate the location of the region and the 4th on indicate its name. The names are of the format PET#\w+.1 for the left side regions and PET#\w+.2 for the right side ones.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

ChiapetExperimentData, loadTFBS, loadPETs, loadPPI, createIndexes.

Examples

```
## get interactions file location
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
res <- CreateCenteredBED(petFile, header=TRUE, dist=1000)
head(res)</pre>
```

createIndexes-methods Preparing TF indexes per region

Description

After loading the interactions and the TFBS, the createIndexes method can be used to build indexes for fast look-up for which which TF are located in which region. This method is an intermediate step needed for further analysis.

Usage

```
## S4 method for signature 'ChiapetExperimentData'
createIndexes(object, minOverlap = 50)
```

16 createServer-methods

Arguments

object a ChiapetExperimentData object in which the interactions and TFBS are al-

ready loaded. Check loadPETs and loadTFBS for more info.

minOverlap The minimum overlap between a TF binding site and a region, to consider a TF

as binding to that region. The default value is 50.

Value

A ChiapetExperimentData object in which the .dt slot is populated as a data.table object.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

ChiapetExperimentData, loadTFBS, loadPETs, loadPPI

Examples

```
## get the different datasets path
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:

x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
## build the diffrent indexes
x <- createIndexes(x)
x

## End(Not run)</pre>
```

createServer-methods

Explore results in a web browser

Description

To facilitate the interaction of the user with the package, we added an additional web interface using the **shiny** package. The user can check some basic statistics about the row data and he can explore the results and generate some graphs.

createServer-methods 17

Usage

```
## S4 method for signature
## 'ChiapetExperimentData,NetworkCollection,ChromMaintainers'
createServer(x,nets,hlda)
```

Arguments

x a ChiapetExperimentData object in which the interactions, TFBS and the in-

dex tables are already created.

nets a NetworkCollection object containing the list of the used TF and the initial

interaction networks.

hlda a ChromMaintainers object in which the results are already calculated.

Value

A webpage is opened.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

NetworkCollection, ChiapetExperimentData, ChromMaintainers

```
## get the different datasets path
 petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")</pre>
 tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")
## Not run:
 data(RPKMS)
 x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
 ## build the diffrent indexes
 x <- createIndexes(x)</pre>
 ## build networks connecting each interacting regions
 nets<- buildNetworks(x)</pre>
 ## infer the networks and do the clustering
 hlda<- InferNetworks(nets)</pre>
 hlda<- clusterInteractions(hlda)</pre>
 ## Run the server
 createServer(x, nets, hlda)
## End(Not run)
```

18 EnsemblToHGNC

EnsemblToHGNC

Ensemble to HGNC conversion

Description

This helper method uses the biomaRt package to convert Ensembl ids to HGNC ids.

Usage

```
EnsemblToHGNC(EnsemblIDs)
```

Arguments

EnsemblIDs a character vector with Ensembl IDs.

Value

returns a data. frame containing the Ensembl ID and his corresponding HGNC gene id and Name plus a description of the gene.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

EntrezToHGNC

EntrezToHGNC 19

EntrezToHGNC

Entrez to HGNC conversion

Description

This helper method uses the biomaRt package to convert Entrez ids to HGNC icS.

Usage

```
EntrezToHGNC(EntrezID)
```

Arguments

EntrezID

a character vector with Entrez IDs.

Value

returns a data.frame containing the Entrez ID and his corresponding HGNC gene id and Name plus a description of the gene.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

EnsemblToHGNC

geneLocations

Nucleus located genes

Description

This dataset contains a data.frame containing the set of genes that are located in the nucleus "GO:0005634".

Usage

```
data(geneLocations)
```

Value

data. frame containing genes located at the nucleus.

Examples

```
data(geneLocations)
head(geneLocations.nucleus)
```

GenerateNetworks-methods

Generate a list of igraph networks

Description

This methods converts the networks slot of a ChromMaintainers object, it reads the topEdge slot and convert it into a list of igraph objects.

Usage

```
## S4 method for signature 'ChromMaintainers'
GenerateNetworks(object,...)
```

Arguments

 $object \hspace{1cm} a \hspace{1cm} ChromMaintainers \hspace{1cm} object \hspace{1cm}$

... future options not considered for the moment.

Value

Returns ChromMaintainers object in which the networks slot is populated.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

ChromMaintainers, InferNetworks

Examples

```
## get the different datasets path
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:
    data(RPKMS)

x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
## build the diffrent indexes
x <- createIndexes(x)
## build networks connecting each interacting regions
nets<- buildNetworks(x)

## infer the networks
hlda<- InferNetworks(nets)

hlda <- GenerateNetworks(hlda)
networks(hlda)

## End(Not run)</pre>
```

getRegionsIncluster-methods

list of interactions per cluster

Description

This method can be used to retrieve the genomic coordinated of the DNA-interactions in each cluster.

Usage

```
## S4 method for signature 'ChromMaintainers,ChiapetExperimentData,numeric'
getRegionsIncluster(hdaRes,data, cluster=1, ...)
```

Arguments

hdaRes	a ChromMaintainers object in which the clusters are already calculated
data	a $\mbox{ChiapetExperimentData}$ object that contains the genomic location of all the interactions.
cluster	The ID of the cluster for which we want to get the list of the involved regions.
	additional parameters not used for the moment.

Value

```
a GRanges object is returned
```

Author(s)

```
Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)
```

See Also

clusterInteractions, InferNetworks, ChiapetExperimentData, ChromMaintainers

```
## get the different datasets path
   petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")</pre>
   tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")</pre>
## Not run:
   data(RPKMS)
   x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
   ## build the different indexes
    x <- createIndexes(x)</pre>
    ## build networks connecting each interacting regions
   nets<- buildNetworks(x)</pre>
    ## infer the networks and do the clustering
   hlda<- InferNetworks(nets)</pre>
   hlda<- clusterInteractions(hlda)</pre>
    ## return the DNA-interactions in cluster 3
    getRegionsIncluster(hlda,x,cluster=3)
## End(Not run)
```

```
getRegionsInNetwork-methods
```

list of interactions per network

Description

This method can be used to retrieve the genomic coordinated of the DNA-interactions enriched for each network given a certain threshold.

Usage

```
## S4 method for signature 'ChromMaintainers, ChiapetExperimentData, numeric'
getRegionsInNetwork(hdaRes, data, net=1, thr=0.5, ...)
```

Arguments

hdaRes	a ChromMaintainers object which already contains the calculated results
data	a ChiapetExperimentData object that contains the genomic location of all the interactions.
net	The ID of the network for which we want to get the list of the involved regions.
thr	Specify the minimum enrichment value to consider an interaction to be controlled by the network. by default it is 0.5
	additional parameters not used for the moment.

Value

```
a GRanges object is returned
```

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

InferNetworks, ChiapetExperimentData, ChromMaintainers

```
## get the different datasets path
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
    tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:
    data(RPKMS)
    x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
    ## build the different indexes
    x <- createIndexes(x)</pre>
```

24 GOEnrich.folder-methods

```
## build networks connecting each interacting regions
nets<- buildNetworks(x)

## infer the networks and do the clustering
hlda<- InferNetworks(nets)

## return the DNA-interactions in cluster 3
getRegionsIncluster(hlda,x,net=3)

## End(Not run)</pre>
```

GOEnrich.folder-methods

GO enrichment methods

Description

This helper methods can be called to do GO enrichment by using the DAVID web service.

GOEnrich. networks can be used to do a GO enrichment of the chromatin maintainer networks.

GOEnrich.folder can be called to do a GO enrichment on the gene-list files generated by the method outputGenesPerClusterToDir.

There is a 5 secs delay between each request to not avoid being rejected by the server.

Usage

```
## S4 method for signature 'character'
GOEnrich.folder(folder, fdr=0.05,GOlimit=20)
## S4 method for signature 'ChromMaintainers'
GOEnrich.networks(object, fdr=0.05, GOlimit= 5,path="")
```

Arguments

folder	name of the folder that contains the gene-list files. The files are supposed to have a .txt extension. The first column of each file is supposed to contain the genes EntrezID.
object	a "ChromMaintainers" objects with the topNodes already calculated.
fdr	cut-off value GO terms with fdr value <= fdr will be considered. Benjamini-Hochberg FDR is used.
GOlimit	the number of top GO terms to return.
path	the path where to store the generated plot (<i>pdf file</i>). if not specified the plot will be displayed.

GOEnrich.folder-methods 25

Value

Returns a list of data. frame that contain the GO results for each file (or network).

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

```
http://david.abcc.ncifcrf.gov/(DAVID website)
```

Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nature Protoc. 2009;4(1):44-57.

See Also

outputGenesPerClusterToDir

```
## get the different datasets path
   petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")</pre>
    tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")</pre>
 ## Not run:
   x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
   ## build the different indexes
   x <- createIndexes(x)</pre>
    ## build networks connecting each interacting regions
   nets<- buildNetworks(x)</pre>
    ## infer the networks
   hlda<- InferNetworks(nets)</pre>
    ## Get the list of genes in each cluster by default
    ## a folder ClustersGenes will be created
    outputGenesPerClusterToDir(hlda,x)
    ## GO enrichment
    GOEnrich.folder(folder="ClustersGenes/")
## End(Not run)
```

26 HLDAResult-class

HLDAResult-class

Class "HLDAResult"

Description

This class is a container for the results generated by the HLDA algorithm

Usage

HLDAResult(docPerTopic,wordsPerTopic ,betas)

Arguments

docPerTopic Object of class "matrix" describing the partnership of each document in a topic.

In our case a documents is the protein interaction networks associated with each DNA interaction and the topics are the inferred chromatin maintainer networks.

wordsPerTopic Object of class "matrix" describing the partnership of each word to a topic. In

our case a word is an edge and the topics are the chromatin maintainer networks.

betas Object of class "numeric" the beta values for each network. Basically, it shows

how popular is the network.

Value

an HLDAResult object.

Accessors

For a given HLDAResult object the following accesor functions can be used:

docPerTopic(x) gets the content of the docPerTopic matrix.

wordsPerTopic(x) gets the content of the wordsPerTopic matrix.

betas(x) gets the betas values.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Chong Wang, John Paisley and David M. Blei, *Online variational inference for the hierarchical Dirichlet process*. In AISTATS 2011

Mohamed Nadhir D, Yang C et al, 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

NetworkCollection, ChromMaintainers, InferNetworks

HPRD 27

Examples

```
showClass("HLDAResult")
```

HPRD

HPRD protein interaction Network

Description

loads an igraph object that contains the HRPD PPI release 9.

Usage

```
data(HPRD)
```

Value

```
an igraph object named PPI. HPRD.
```

Source

http://hprd.org/

Examples

```
data(HPRD)
PPI.HPRD
```

InferNetworks-methods Network construction using Hierarchical Dirichlet Process

Description

This methods applies a Hierarchical Dirichlet Process (HDP) algorithm on the collection of proteins networks to infer the set of chromatin loop-maintainer proteins. HDP are non-parametric Bayesian models widely used in document classification as it enables us to model datasets with a mixtures of classes. In our case, we suppose that different kinds of networks are involved in maintaining the different loops. Thus, to make an analogy with topic modeling, we each DNA-interaction maintaining protein network as a document and each edge in this network as word. Thus, the task is to say which word (*edge*) belongs to which topic (*chromatin-maintainer family*). The method implementation is based on the C++ code of Chong Wang and David Blei with adaptation to Rcpp and removal of the dependency on the Gnu Scientific Library.

Usage

```
## S4 method for signature 'NetworkCollection'
InferNetworks(object,thr =0.5,max_iter = 500L, max_time = 3600L, ...)
```

28 InferNetworks-methods

Arguments

a NetworkCollection object in which the list of protein interactions associated with each DNA interaction is already populated.

thr

Used to select the top protein interaction in each inferred chromatin-maintainer family. In HDP each topic (*Chromatin-maintainer family*) is considered as a distribution over words (*edges*), thus, for each topic we consider the words that capture threshold percent of the topic to be the top words. For example, in topic1, we first rank edges by partnership probability to topic1 in a decreasing order, and we take the top edges that capture 50% of the partnership.

max_iter maximum number of iterations (befault 500).

max_time maximum runing time (3600 sec).

... not used for the moment, left for additional parameters.

Value

Returns a ChromMaintainers object that contains the list of inferred networks and the probability of each edge in each network.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

```
https://www.cs.princeton.edu/~blei/topicmodeling.html (C. Wang's hdp code)
```

Chong Wang, John Paisley and David M. Blei, *Online variational inference for the hierarchical Dirichlet process*. In AISTATS 2011

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

NetworkCollection, ChromMaintainers

```
## get the different datasets path
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
    tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:

x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
    ## build the different indexes
    x <- createIndexes(x)
    ## build networks connecting each interacting regions
    nets<- buildNetworks(x)

## infer the networks
hlda<- InferNetworks(nets)</pre>
```

loadPETs-methods 29

hlda

End(Not run)

loadPETs-methods

Parsing ChIA-PET interaction data

Description

This method loads the ChIA-PET interactions data into a GRanges object from a file generated by ChIA-PET tool or an already formatted BED file.

Usage

```
## S4 method for signature 'ChiapetExperimentData,character'
loadPETs(object, petFile, IsBed=TRUE, header=TRUE, dist=1000)
```

Arguments

object

(Required) a ChiapetExperimentData object

petFile

(**Required**) path to the file to parse. two types of formats are accepted. If the file comes from the ChIA-PET tool or has the same format used in the ENCODE project

```
chromleft startleft endleft chromright startright endright
    chr1
            872113 879175
                                  chr1
                                           933836
                                                     938416
    chr1
            874165 879175
                                  chr1
                                           933340
                                                     938306
    chr1
            889676 896594
                                  chr1
                                           933897
                                                     938982
    chr1
            898753 907581
                                  chr1
                                           931133
                                                     939571
    chr1
            910103
                    918775
                                  chr1
                                           930834
                                                     938627
    chr1
            919314 922154
                                  chr1
                                           934212
                                                     937864
```

only the the coordinates of the left and the right regions are considered (which means the first 6 columns), The additional columns are just ignored. We suppose that the user has already selected the interactions that sound significant for him. if this kind of file is provided the IsBed parameter should be set to FASLE. if the file has no header the user needs to set header=FALSE.

if IsBed parameter is provided, the provided file should have 4 columns, three to describe the region location and the forth column to indicate the ID of the interaction and its position (1: for left and 2: for right), according to following pattern PET#\d\.1 or PET#\d\.2.

```
chr1
       1241234
                  1242234 PET#1.1
chr1
       1242724
                  1243724 PET#1.2
chr1
       1282708
                 1283708 PET#2.1
       1283834
                 1284834 PET#2.2
chr1
       1370871
                  1371871 PET#3.1
chr1
       1372322
                 1373322 PET#3.2
chr1
```

30 loadPETs-methods

IsBed (optional) The flag indicates whether the provided file has a 4 columns BED

(when IsBed = TRUE) format or has the ChIA-PET tool format (when IsBed = FALSE).

By default, IsBed = TRUE

header (optional) Indicates whether or not the first line of the file should be considered

as a header. by default it is TRUE

dist (optional) This parameter indicates the size of the up- and down-stream regions

(in bp) to consider around the center of each region. 3CPET is based on the assumption that the real ChIP-Seq signal is more enriched around the center of the regions and get more depleted when moving further in both directions. Thus, when the user provides a file that has a ChIA-PET tool format (IsBed = TRUE), 3CPET will first center the regions and consider only distance bp up- and down-stream of the center of each region. A default value of 1000bp is used.

Value

A ChiapetExperimentData object in which the ppi is populated as a GRanges object.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Li G, Fullwood MJ, Xu H et al. ChIA-PET tool for comprehensive chromatin interaction analysis with paired-end tag sequencing. Genome Biology 2010, 11(2):R22

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

ChiapetExperimentData, loadTFBS, loadPPI, createIndexes

```
## Create a ChiapetExperimentData object
x <- ChiapetExperimentData(ppiType= "HPRD")

## load the different datasets (where the file has a Chia-PET tool format )
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
x <- loadPETs(x, petFile=petFile, IsBed=FALSE)

## when loading an already formatted BED file
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_centered.bed")
x <- loadPETs(x, petFile=petFile, IsBed=TRUE, header=FALSE)

pet(x)</pre>
```

loadPPI-methods 31

loadPPI-methods	Setting the background protein-protein interaction	

Description

This method enables the user to define the PPI network as a background network. The user can provide his own PPI or use the HPRD or the Biogrid PPI incorporated in the package.

Usage

```
## S4 method for signature 'ChiapetExperimentData'
loadPPI(object,type=c("HPRD","Biogid"),customPPI= NULL,
              filter=FALSE, term ="GO:0005634", annot=NULL,
              RPKM= NULL, threshold=1 )
```

Arguments

term

annot

object a ChiapetExperimentData to load the PPI into as an igraph object

type if customPPI = NULL, this parameter indicats which of the available PPI to use

HPRD or Biogrid (build 3.2.100). by default, the HPRD network is used.

customPPI If the user wants to use his own PPI interaction network (for example for another

> species), he can provide an igraph object or a path to an "ncol" formatted graph (a file with two columns indicating the interacting parts and an optional third column to indicate the weight). It is preferred that the user provides an

igraph object, to avoid any problems when parsing the "ncol" file.

filter This parameter indicates whether the user want to filter the selected PPI or not.

> filter = FALSE means that the PPI (provided by the user or by the HPRD and Biogrid networks) will be used as is. if filter = TRUE, the proteins in the PPI will be filtered according to their position in cell (by default proteins located in

the nucleus are kept and the other removed).

In addition to the filtering by location, if the user wants just to keep the proteins that show a certain amount of expression he can provide a gene expression table

to the RPKM parameter and set the threshold for filtering.

The GO term used to for filtering. By default, only protein that are located in the nucleus are kept (term = "GO:0005634"). If the user want to use another

annotation table he can pass a data. frame object to the parameter annot.

If the user wants to provide his own annotation data-set, he can pass a data. frame object to this parameter. The passed data.frame should have at least two columns named respectively: geneSymbol that contains the gene names, and

cellular_component_term that contain the term.

```
geneSymbol
                               cellular_component_term
FAU Ribosome
                  (GO:0005840); Nucleolus (GO:0005730)
ALDH3A1
          Cytoplasm (GO:0005737); Nucleus (GO:0005634)
ASCL1
          Nucleus (GO:0005634); Cytoplasm (GO:0005737)
```

32 loadPPI-methods

RPKM A data. frame that contains the expression of each gene in the PPI. The data. frame

should at least have 2 columns. The first one contains the gene symbol (should be the same as the one used in the PPI) and the second gives the expression.

threshold Threshold value used to filter gene expression. All genes with expression value

less than threshold are removed.

Value

A ChiapetExperimentData object in which the ppi slot is populated as an igraph object filtered according to the specified conditions.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Prasad, T. S. K. et al. (2009) *Human Protein Reference Database - 2009 Update*. Nucleic Acids Research. 37, D767-72.

Chatr-Aryamontri A, Breitkreutz BJ et al. *The BioGRID Interaction Database: 2013 update*. Nucleic Acids Res. 2012 Nov 30

M.N Djekidel et al, 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process, in press, 2015

See Also

ChiapetExperimentData, loadTFBS, loadPETs, createIndexes

```
## Create a ChiapetExperimentData object
x <- ChiapetExperimentData(ppiType= "HPRD")

## Loading the default HPRD network without filtering

x <- loadPPI(x,type="HPRD")
ppi(x)

## Using the HPRD network and filtering using the GO:0005634
x <- loadPPI(x,type="HPRD", filter=TRUE)
ppi(x)

data(RPKMS)
x <- loadPPI(x,type="HPRD",filter=TRUE,annot= NULL, RPKM= RPKMS, threshold = 5)
ppi(x)</pre>
```

loadTFBS-methods 33

Description

This methods reads a BED file that contains the peak positions of different TF. All the TF peaks should be merged into one BED file that contains 4 columns that respectively contain the chromosome name, peak start, peak end, TF name.

Usage

```
## S4 method for signature 'ChiapetExperimentData,character'
loadTFBS(object, tfbsFile,header=FALSE, ...)
```

Arguments

guments	
object	(Required) a ChiapetExperimentData object
tfbsFile	(Required): path the BED file containing the position of the different TF binding site All the TF binding sites should be merged in this file as showed in this example:
	chr1 569820 569998 BHLHE40
	chr1 936071 936346 BHLHE40
	chr1 1014795 1015082 BHLHE40
	chrY 13485240 13485769 ZBTB33
	chrY 13488718 13489030 ZBTB33
	chrY 15016340 15016848 ZBTB33
	chrY 58843918 58844104 ZBTB33
h a a da u	(autional) indicates if the annuited DED 61s has a harden or made but default
neader	(optional) indicates if the provided BED file has a header or node. by default header=FALSE,
	reserved for later use.
	object tfbsFile header

Value

A ChiapetExperimentData object in which the tfbs slot is populated as a GRanges object.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

34 NetworkCollection-class

See Also

ChiapetExperimentData, loadTFBS, loadPPI, createIndexes

Examples

```
## Create a ChiapetExperimentData object
x <- ChiapetExperimentData(ppiType= "HPRD")

## load TFBS
tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")
x <- loadTFBS(x,tfbsFile=tfbsFile)
tfbs(x)</pre>
```

NetworkCollection-class

protein interaction networks maintaining DNA loops

Description

The class NetworkCollection stores information about the set of protein networks that maintains DNA interactions.

Usage

```
NetworkCollection(networks, sizes, TFCollection)
```

Arguments

networks a list of list, each list contains the set of edges in each network

sizes the sizes of each network. The should correspond to the sizes of the networks

TFCollection the set of all the TF involved in all the interactions.

Details

The NetworkCollection contains three main information: (i) the set of edges in each network maintaining each DNA loop, (ii) the number of edges in each network and (iii) the set of TF involved in all the networks.

Value

a NetworkCollection object.

Accesors

```
networks gets the list of networkssizes gets the vector containing the size of each networkTF gets the list of involved TF (after filtering)
```

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

See Also

InferNetworks, ChiapetExperimentData, buildNetworks

Examples

```
showClass("NetworkCollection")
```

outputGenesPerClusterToDir-methods

List of genes in each cluster

Description

This helper methods get the set of genes located in the DNA-regions in each cluster. A folder that contain a bunch of .txt files (one for each cluster) is generated. We consider a gene to be part of a cluster if the (-1500bp, +500bp) around its TSS intersects with one of the DNA regions of the cluster.

Usage

```
## S4 method for signature 'ChromMaintainers, ChiapetExperimentData' outputGenesPerClusterToDir(hdaRes, data, path="ClustersGenes", ...)
```

Arguments

hdaRes	a ${\sf ChromMaintainers}$ object in which the DNA-interaction are already clustered.
data	a ${\tt ChiapetExperimentData}$ object containing information about the interactions.
path	path of the folder to create. by default a folder named ClustersGenes is created in the current working directory.
	additional parameters, not used for the moment.

Value

The specified folder is created with a list .txt files that contain the list of genes.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

ChromMaintainers, InferNetworks, ChiapetExperimentData, clusterInteractions

Examples

```
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
   tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:
   x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
   ## build the different indexes
   x <- createIndexes(x)

   ## build networks connecting each interacting regions
   nets<- buildNetworks(x)

   ## infer the networks
   hlda<- InferNetworks(nets)
   hlda<- clusterInteractions(hlda)

   ## get the list of genes per cluster.
   outputGenesPerClusterToDir(hlda,x)

## End(Not run)</pre>
```

Description

outputGenesPerNetworkToDir

This helper methods get the set of genes located in the DNA-regions controlled by each network. A folder that contains a bunch of .txt files (one for each network) is generated. We consider (-2500bp, +2500bp) around the TSS of gene located in a region showing 0.5 or more enrichment for the network.

List of genes controlled by each network

Usage

```
## S4 method for signature 'ChromMaintainers, ChiapetExperimentData' outputGenesPerNetworkToDir(hdaRes, data, path="NetworksGenes", ...)
```

Arguments

hdaRes	a ChromMaintainers object containing the enrichment values.
data	a $\mbox{ChiapetExperimentData}$ object containing information about the interactions.
path	path of the folder to create. by default a folder named NetworksGenes is created in the current working directory.
	additional parameters, not used for the moment.

Value

The specified folder is created with a list .txt files each for each network that contain the list of genes.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

ChromMaintainers, InferNetworks, ChiapetExperimentData

```
## get the different datasets path
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")
## Not run:
    data(RPKMS)

x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
    ## build the different indexes
    x <- createIndexes(x)

## build networks connecting each interacting regions
nets<- buildNetworks(x)

## infer the networks
hlda<- InferNetworks(nets)

## get the list of genes per network.
outputGenesPerNetworkToDir(hlda,x)

## End(Not run)</pre>
```

plot3CPETRes-methods

Plotting clustering results

Description

This method enables the user the generate different types of plots to visualize the results.

Usage

Arguments

object (Required) a ChromMaintainers object that contains the results

(optional) path where to save the plots should be ".pdf". if not provided the plot will be displayed on the screen.

(optional) The width of the plot in the pdf file.by default it is 14 inch

(optional) The Height of the plot in the pdf file.by default it is 7 inch

type type of the plot to generate. It can support the following values ((default: "heatmap")):

- "heatmap": Generates a heatmap of the partnership of each DNA interaction to a chromatin maintainer network. Each column is a DNA interaction and each row a chromatin maintainer network.
- "clusters": generates a pair-wise scatter plots of all clusters. *Note:* only supported if the sota method was applied.
- "curve": for each cluster plots the enrichment profile of all the elements.
- "avgCurve": draws the average curve of the enrichment profile for each clusters.
- "netSim": plots a heatmap showing the percentage of common proteins or
 edges between the chromatin maintainer networks. *Note:* generally the similarity between networks is so small so the user can set the value in the diagonal to zero and then re-plot or plot the dissimilarity plot. if byEdge = TRUE
 a similarity based on the common edges is calculated otherwise by common
 nodes.
- "networks": plots all the networks. if this option is chosen a pdf file named "AllGraphs.pdf" is generated in the current working unless the path parameter is explicitly determined. To get a finer control, the user can specify the type of layout to use, by default the layout.kamada.kawai is used. For additional layout functions you can check the igraph package. The reason we generate a pdf file because there is a lot of networks and it will not be convenient to display them in one plot, or generating multiple plots.

byEdge (optional) if TRUE and type = "netSim" then the a heatmap showing the similarity between the chromatin maintainer networks by common edges. if FALSE the similarity is calculated based on the number of common nodes.

1 (optional) The graph layout algorithm to use. by default the layout.kamada.kawai is used. Additional functions are available in the igraph package.

options for future use.

Value

Different types of values are returned depending on the type of the plot selected.

"heatmap" returns a list generated by the pheatmap method, however it is always empty.

"clusters", "curve", "avgCurve" returns a list describing the number of plots per row and col-

"netSim" returns a list that contains a ggplot2 object and the similarity matrix

"networks" returns a list of ggplot2objects, one per network.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

```
cluster, igraph, sota
```

```
## get the different datasets path
   petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")</pre>
   tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")
## Not run:
   data(RPKMS)
   x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
   ## build the different indexes
   x <- createIndexes(x)</pre>
   ## build networks connecting each interacting regions
   nets<- buildNetworks(x)</pre>
    ## infer the networks
   hlda<- InferNetworks(nets)</pre>
    ## cluster results
   hlda<- clusterInteractions(hlda)</pre>
    ## plot a heatmap
   plot3CPETRes(hlda, type="heatmap")
```

40 plotTrack-methods

```
## plot clusters pair-wise scatter plots
plot3CPETRes(hlda, type="clusters")

## enrichment plot for the elements in each network
plot3CPETRes(hlda, type="curve")

## average enrichment plot for the elements in each network
plot3CPETRes(hlda, type="avgCurve")

## heatmap showing the similarity between the different network
plot3CPETRes(hlda, type="netSim")

## plot all the networks in the file "AllGraphs.pdf"
nets_plot <- plot3CPETRes(hlda, type="networks")

## plot one of the networks
plot(nets_plot[[3]])

## End(Not run)</pre>
```

plotTrack-methods

Plot interaction on a genomic track

Description

This helper method can be used to display a genomic track for a certain location that contains the chromosome and the related interactions if any.

Usage

```
## S4 method for signature 'ChiapetExperimentData,GRanges'
plotTrack(object, range)
```

Arguments

object a ChiapetExperimentData object which contains the raw data.

range The genomic coordinates of the location to display

Value

```
a ggbio::track object
```

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

ChiapetExperimentData

PrepareData-methods 41

Examples

```
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
    tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:
    x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
    gr <- GRanges("chr1",IRanges(start=100000,end=300000))
    plotTrack(x,gr)

## End(Not run)</pre>
```

PrepareData-methods

Loading the raw data all at once

Description

Instead of loading the data one at a time and then creating the index using the methods loadPETs,loadTFBS and createIndexes. The user can directly use the method PrepareData to do that.

Usage

```
## S4 method for signature 'character,character,logical'
PrepareData(petFile,tfbsFile, petIsBed=TRUE)
```

Arguments

petFile	a character specifying the path to the interaction file. it the file is in a "bed" format petIsBed should be TRUE. The data should be formated as described in loadPETs.
tfbsFile	a character specifying the path to the transcription factors binding site file. The data should be formated as described in loadTFBS.
petIsBed	a logical value specifying if the interaction file is in a "bed" format or not.

Value

A ChiapetExperimentData object in which the pet,tfbs and .dt slots populated .

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

References

Mohamed Nadhir D, Yang C et al 3CPET: Finding Co-factor Complexes in Chia-PET experiment using a Hierarchical Dirichlet Process,

42 RPKMS

See Also

ChiapetExperimentData, loadTFBS, loadPETs, loadPPI, createIndexes.

Examples

```
## get interactions file location
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")

## get the TFBS file location
tfFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:
## load the data
x<- PrepareData(petFile, tfFile, FALSE)
x

## End(Not run)</pre>
```

RPKMS

A gene expression dataset

Description

A gene expression dataset for the K562 cell line.

Usage

```
data(RPKMS)
```

Value

a data. frame containing genes expression in K562 cells.

```
data(RPKMS)
head(RPKMS)
```

updateResults-methods 43

updateResults-methods update top maintain networks elements

Description

This helper method can be used to update the list of interactions constituting the chromatin maintainer networks by changing the threshold.

Usage

```
## S4 method for signature 'ChromMaintainers,NetworkCollection,numeric'
updateResults(object,nets,thr=0.5)
```

Arguments

object a ChromMaintainers object in which the clusters are already calculated

nets a NetworkCollection used to get the vocabulary list.

thr The cut-off threshold, the top edges that have an enrichment value that sum up

to a value >= thr are considered.

Value

a ChromMaintainers object in which the topNodes and topEdges tables are updated.

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

InferNetworks, NetworkCollection, ChromMaintainers

```
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
    tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:
    data(RPKMS)

x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
    ## build the different indexes
    x <- createIndexes(x)

## build networks connecting each interacting regions
    nets<- buildNetworks(x)

## infer the networks</pre>
```

44 visualizeCircos-methods

```
hlda<- InferNetworks(nets)
topNodes(hlda)

hlda <- updateResults(hlda, nets, 0.4)
topNodes(hlda)

## End(Not run)</pre>
```

visualizeCircos-methods

Generate circos plot per cluster

Description

This method generates a basic circos plot of the chromatin interaction in a given cluster.

Usage

```
## S4 method for signature 'ChromMaintainers, ChiapetExperimentData, numeric'
visualizeCircos(object, data, cluster = 1, chrLenghts = NULL)
```

Arguments

object a ChromMaintainers object in which the clusters are already calculated.

data a ChiapetExperimentData containing the interaction data.

cluster the number of the cluster to display

chrLenghts the chromatin lengths. if not provided the package suppose it is a human chro-

matin and uses the corresponding lengths. Change it if you are using another

species.

Value

circos a GRanges object that contains the coordinate of the left side interactions. The right side interactions can be accessed by writing circos\$to.gr.

```
plot a ggplot object
```

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

NetworkCollection, ChromMaintainers

visualizeInteractions 45

Examples

```
## get the different datasets path
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
    tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:
    data(RPKMS)
    x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
    ## build the different indexes
    x <- createIndexes(x)

## build networks connecting each interacting regions
    nets<- buildNetworks(x)

## infer the networks
    hlda<- InferNetworks(nets)
    hlda<- clusterInteractions(hlda)
    visualizeCircos(hlda,x, cluster=3)

## End(Not run)</pre>
```

Description

This method can be used to draw a circos plot of the chromatin interactions located in the given genomic range.

Usage

```
## S4 method for signature 'ChiapetExperimentData, GRanges'
visualizeInteractions(object, range)
```

Arguments

object a ChiapetExperimentData in which the interactions are already loaded. Check

loadPETs for more info.

range a GRanges object containing the genomic region of interest.

Value

A ciros plot of the selected region is displayed and a list containing the following objects is returned.

```
circos: a GRanges object that contains the involved chromatin interactions.
```

plot: a ggplot object containing plot.

46 visualizeInteractions

Author(s)

Mohamed Nadhir Djekidel (<nde12@mails.tsinghua.edu.cn>)

See Also

ChiapetExperimentData, loadPETs, ggbio, GRanges

```
petFile <- file.path(system.file("example",package="R3CPET"),"HepG2_interactions.txt")
tfbsFile <- file.path(system.file("example",package="R3CPET"),"HepG2_TF.txt.gz")

## Not run:

x <- ChiapetExperimentData(pet = petFile, tfbs= tfbsFile, IsBed = FALSE, ppiType="HPRD", filter= TRUE)
## plot intractions in the region of interest
gr <- GRanges("chr1", IRanges(1240000,10300000))
p <- visualizeInteractions(x, gr)
p

## End(Not run)</pre>
```

Index

*Topic \textasciitilde\textasciitilde	${\tt getRegionsInNetwork-methods}, {\tt 23}$
other possible keyword(s)	GOEnrich.folder-methods, 24
\textasciitilde\textasciitilde	InferNetworks-methods, 27
annotate Expression-methods, 3	loadPETs-methods, 29
clusterInteractions-methods, 13	loadPPI-methods, 31
createIndexes-methods, 15	loadTFBS-methods, 33
createServer-methods, 16	$output {\tt GenesPerClusterToDir-methods},$
GenerateNetworks-methods, 20	35
<pre>getRegionsIncluster-methods, 21</pre>	outputGenesPerNetworkToDir, 36
<pre>getRegionsInNetwork-methods, 23</pre>	plot3CPETRes-methods, 38
InferNetworks-methods, 27	plotTrack-methods, 40
loadPPI-methods, 31	PrepareData-methods, 41
plot3CPETRes-methods, 38	updateResults-methods, 43
plotTrack-methods, 40	visualizeCircos-methods, 44
PrepareData-methods, 41	visualizeInteractions, 45
updateResults-methods, 43	*Topic package
*Topic classes	R3CPET-package, 2
ChiapetExperimentData-class, 6	
ChromMaintainers-class, 10	annotateExpression
cluesOrSota-class, 12	(annotateExpression-methods), 3
HLDAResult-class, 26	annotateExpression,ChromMaintainers,data.frame-method
NetworkCollection-class, 34	(annotateExpression-methods), 3
*Topic datasets	${\tt annotate Expression-methods}, 3$
Biogrid, 4	
chromosoms, 12	betas (HLDAResult-class), 26
geneLocations, 20	betas, HLDAResult-method
HPRD, 27	(HLDAResult-class), 26
RPKMS, 42	betas-methods (HLDAResult-class), 26
*Topic methods	Biogrid, 4
annotate Expression-methods, 3	buildNetworks, 5, 35
buildNetworks, 5	<pre>buildNetworks,ChiapetExperimentData-method</pre>
clusterInteractions-methods, 13	(buildNetworks), 5
CreateCenteredBED-methods, 14	
<pre>createIndexes-methods, 15</pre>	ChiapetExperimentData, 3, 5, 6, 8, 15-17,
createServer-methods, 16	22, 23, 29–37, 40–42, 44–46
EnsemblToHGNC, 18	ChiapetExperimentData
EntrezToHGNC, 19	(ChiapetExperimentData-class),
GenerateNetworks-methods, 20	6
<pre>getRegionsIncluster-methods, 21</pre>	ChiapetExperimentData-class, 6

48 INDEX

ChromMaintainers, <i>3</i> , <i>4</i> , <i>11–14</i> , <i>17</i> , <i>20–23</i> ,	GenerateGmlNetworks,ChromMaintainers-method
26, 28, 35–37, 43, 44	(ChromMaintainers-class), 10
ChromMaintainers	GenerateGmlNetworks-methods
(ChromMaintainers-class), 10	(ChromMaintainers-class), 10
ChromMaintainers-class, 10	GenerateNetworks
Chromosoms (chromosoms), 12	(GenerateNetworks-methods), 20
chromosoms, 12	GenerateNetworks,ChromMaintainers-method
clues, <i>12–14</i>	(GenerateNetworks-methods), 20
cluesOrSota-class, 12	GenerateNetworks-methods, 20
cluster, 39	getClusters (ChromMaintainers-class), 10
clusterInteractions, 22, 36	getClusters,ChromMaintainers-method
clusterInteractions	(ChromMaintainers-class), 10
<pre>(clusterInteractions-methods),</pre>	getClusters-methods
13	(ChromMaintainers-class), 10
clusterInteractions,ChromMaintainers-method	getRegionsIncluster
<pre>(clusterInteractions-methods),</pre>	<pre>(getRegionsIncluster-methods),</pre>
13	21
clusterInteractions-methods, 13	${\tt getRegionsIncluster,ChromMaintainers,ChiapetExperimentData}$
clValid, <i>13</i>	<pre>(getRegionsIncluster-methods),</pre>
CreateCenteredBED	21
(CreateCenteredBED-methods), 14	<pre>getRegionsIncluster-methods, 21</pre>
CreateCenteredBED,character-method	getRegionsInNetwork
(CreateCenteredBED-methods), 14	<pre>(getRegionsInNetwork-methods),</pre>
CreateCenteredBED-methods, 14	23
createIndexes, 6, 8, 15, 30, 32, 34, 41, 42	<pre>getRegionsInNetwork,ChromMaintainers,ChiapetExperimentData</pre>
<pre>createIndexes (createIndexes-methods),</pre>	(getRegionsInNetwork-methods),
15	23
<pre>createIndexes,ChiapetExperimentData-method</pre>	getRegionsInNetwork-methods, 23
(createIndexes-methods), 15	ggbio, 46
createIndexes-methods, 15	ggplot, 44, 45
createServer (createServer-methods), 16	ggnlot2.39
${\sf createServer}$, ${\sf ChiapetExperimentData}$, ${\sf NetworkCommunitarity}$	llection ChromMaintainers-method
(createServer-methods), 16	(GOEnrich.folder-methods), 24
createServer-methods, 16	GOEnrich.folder,character-method
	(GOEnrich.folder-methods), 24
data.frame, <i>18</i> , <i>19</i> , <i>31</i> , <i>32</i>	GOEnrich.folder-methods, 24
data.table, <i>16</i>	GOEnrich.networks
docPerTopic (HLDAResult-class), 26	(GOEnrich.folder-methods), 24
docPerTopic,HLDAResult-method	GOEnrich.networks,ChromMaintainers-method
(HLDAResult-class), 26	(GOEnrich.folder-methods), 24
docPerTopic-methods(HLDAResult-class),	GOEnrich.networks-methods
26	(GOEnrich.folder-methods), 24
- 12- 12- 12- 12	GRanges, 7, 8, 22, 23, 29, 30, 33, 44–46
EnsemblToHGNC, 18, 19	5g55, 7, 6, 22, 26, 27, 66, 66, 77, 76
EntrezToHGNC, 18, 19	UI DADocul+ 2 10 11
7 20	HLDAResult, 3, 10, 11
geneLocations, 7, 20	HLDAResult (HLDAResult-class), 26
GenerateGmlNetworks	HLDAResult-class, 26
(ChromMaintainers-class), 10	HPRD, 27

INDEX 49

igraph, 4, 5, 8, 10, 11, 20, 27, 31, 32, 38, 39	outputGenesPerNetworkToDir-methods
InferNetworks, 4, 10, 11, 14, 21–23, 26, 35–37, 43	<pre>(outputGenesPerNetworkToDir), 36</pre>
•	30
InferNetworks (InferNetworks-methods), 27	<pre>pet (ChiapetExperimentData-class), 6</pre>
	pet,ChiapetExperimentData-method
InferNetworks, NetworkCollection-method	(ChiapetExperimentData-class),
(InferNetworks-methods), 27	6
InferNetworks-methods, 27	pet-methods
loadPETs, 5–9, 15, 16, 32, 41, 42, 45, 46	(ChiapetExperimentData-class),
loadPETs (loadPETs-methods), 29	6
loadPETs,ChiapetExperimentData,character-me	thmet<- (ChianatEvnerimentData-class) 6
(loadPETs-methods), 29	pet<-,ChiapetExperimentData-method
loadPETs-methods, 29	
	(ChiapetExperimentData-class),
loadPPI, 5-9, 15, 16, 30, 34, 42	nhaatman 20
loadPPI (loadPPI-methods), 31	pheatmap, 39
loadPPI, ChiapetExperimentData-method	plot3CPETRes (plot3CPETRes-methods), 38
(loadPPI-methods), 31	plot3CPETRes, ChromMaintainers-method
loadPPI-methods, 31	(plot3CPETRes-methods), 38
loadTFBS, 5, 6, 8, 9, 15, 16, 30, 32, 34, 41, 42	plot3CPETRes-methods, 38
loadTFBS (loadTFBS-methods), 33	plotTrack (plotTrack-methods), 40
loadTFBS,ChiapetExperimentData	plotTrack, ChiapetExperimentData, GRanges-method
(loadTFBS-methods), 33	(plotTrack-methods), 40
loadTFBS,ChiapetExperimentData,character-me	
(loadTFBS-methods), 33	<pre>ppi (ChiapetExperimentData-class), 6</pre>
loadTFBS-methods, 33	ppi,ChiapetExperimentData-method
	(ChiapetExperimentData-class),
NetworkCollection, 6, 17, 26, 28, 43, 44	6
NetworkCollection	ppi-methods
(NetworkCollection-class), 34	(ChiapetExperimentData-class),
NetworkCollection-class, 34	6
networks (NetworkCollection-class), 34	PPI.Biogrid (Biogrid), 4
networks, ChromMaintainers-method	PPI.HPRD (HPRD), 27
(ChromMaintainers-class), 10	<pre>ppi<- (ChiapetExperimentData-class), 6</pre>
networks,NetworkCollection-method	ppi<-,ChiapetExperimentData-method
(NetworkCollection-class), 34	(ChiapetExperimentData-class),
outputGenesPerClusterToDir, 25	6
outputGenesPerClusterToDir	PrepareData (PrepareData-methods), 41
(outputGenesPerClusterToDir-methods)	PrepareData, character, character, logical-method
35	(,,,,,,
outputGenesPerClusterToDir,ChromMaintainers	PrepareData-methods, 41
(outputGenesPerClusterToDir-methods)	
35	
	R3CPET-package, 2
outputGenesPerClusterToDir-methods, 35	RPKMS, 7, 42
outputGenesPerNetworkToDir, 36	Chizant Entweithent Nothing thidan \ 24
outputGenesPerNetworkToDir,ChromMaintainers	
<pre>(outputGenesPerNetworkToDir),</pre>	sizes, NetworkCollection-method
36	(NetworkCollection-class), 34

50 INDEX

```
sota, 12-14, 38, 39
TF (NetworkCollection-class), 34
TF, NetworkCollection-method
        (NetworkCollection-class), 34
tfbs (ChiapetExperimentData-class), 6
tfbs, ChiapetExperimentData-method
        (ChiapetExperimentData-class),
tfbs-methods
        (ChiapetExperimentData-class),
tfbs<- (ChiapetExperimentData-class), 6
tfbs<-,ChiapetExperimentData-method
        (ChiapetExperimentData-class),
topEdges (ChromMaintainers-class), 10
topEdges, ChromMaintainers-method
        (ChromMaintainers-class), 10
topEdges-methods
        (ChromMaintainers-class), 10
topNodes (ChromMaintainers-class), 10
topNodes, ChromMaintainers-method
        (ChromMaintainers-class), 10
topNodes-methods
        (ChromMaintainers-class), 10
updateResults (updateResults-methods),
updateResults, ChromMaintainers, NetworkCollection, numeric-method
        (updateResults-methods), 43
updateResults-methods, 43
visualizeCircos
        (visualizeCircos-methods), 44
visualizeCircos,ChromMaintainers,ChiapetExperimentData,numeric-method
        (visualizeCircos-methods), 44
visualizeCircos-methods, 44
visualizeInteractions, 45
visualizeInteractions, ChiapetExperimentData, GRanges-method
        (visualizeInteractions), 45
visualizeInteractions-methods
        (visualizeInteractions), 45
wordsPerTopic (HLDAResult-class), 26
wordsPerTopic,HLDAResult-method
        (HLDAResult-class), 26
wordsPerTopic-methods
        (HLDAResult-class), 26
```