

Package ‘HiCfeat’

December 4, 2018

Type Package

Title Multiple Logistic Regression for 3D Chromatin Domain Border Analysis

Version 1.4

Depends R (>= 3.2.0), GenomicRanges, Matrix, glmnet, rtracklayer

Imports IRanges, GenomeInfoDb, methods

Date 2018-12-04

Author Raphael Mourad

Maintainer Raphael Mourad <raphael.mourad@ibcg.biotoul.fr>

Description We propose a multiple logistic regression model to assess the influences of genomic features such as DNA-binding proteins and functional elements on topological domain borders.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2018-12-04 11:10:07 UTC

R topics documented:

HiCfeat-package	2
averagePerBin	2
borderAnalysisFun	3
dataExample	5
readDomBed	6
readGFBed	6
readGF Wig	7

Index	9
--------------	----------

HiCfeat-package	<i>Multiple Logistic Regression for 3D Chromatin Domain Border Analysis</i>
-----------------	---

Description

We propose a multiple logistic regression model to assess the influences of genomic features such as DNA-binding proteins and functional elements on topological domain borders.

Author(s)

Raphael Mourad

Maintainer: Raphael Mourad, raphael.mourad@ibcg.biotoul.fr

References

Raphael Mourad and Olivier Cuvier. Computational identification of genomic features that influence 3D chromatin domain formation, PLOS Computational Biology, 12(5):e1004908.

averagePerBin	<i>Compute average value of a genomic feature for each bin.</i>
---------------	---

Description

This function is for internal use only! The function computes the average value of a genomic feature for each bin of size binsize.

Usage

```
averagePerBin(x, binsize, mcolname)
```

Arguments

x	The GRanges object used to compute the average for each bin.
binsize	Size of each bin.
mcolname	The metadata column name used to compute the average for each bin.

Value

A vector object is returned.

Author(s)

Raphael Mourad

borderAnalysisFun *Multivariate enrichment test for chromatin domain borders*

Description

This is the main function. This function estimates the influences of genomic features on chromatin domain borders by a multiple logistic regression model. Enrichment test by simple logistic regression is also included in this function for comparisons. Genomic features can be either coordinate data (e.g. ChIP-seq peak or functional element coordinates) or quantitative data (e.g. normalized ChIP-seq values: $\log_2(\text{ChIP}/\text{Input})$ values). The proposed method is flexible and can account for statistical interactions among multiple genomic features.

Usage

```
borderAnalysisFun(genomicFeatureList.GR, GFDataType, annotNames, domains.GR,
  seqInfoChr, analysisMode, binSize, borderSize,
  LRT, interactionTerms, verbose)
```

Arguments

genomicFeatureList.GR	A list of GRanges objects. Each GRanges object has been built from either coordinate data using readGFBed function (for instance ChIP-seq peak coordinates or functional element coordinates), or quantitative data using readGF Wig (e.g. normalized ChIP-seq values: $\log_2(\text{ChIP}/\text{Input})$ values). All GRanges object should be either coordinate data or quantitative data. There cannot be a mix of GRanges from coordinate data and quantitative data.
GFDataType	GFDataType = "bed" if all GRanges were built from readGFBed function, or GFDataType = "wig" if all GRanges were built from readGF Wig function.
annotNames	A character vector defining the name of each genomic feature. Names should not comprise any special character such as ":+-*^,;!?" because an internal R formula object is created inside the function borderAnalysisFun.
domains.GR	A GRanges for the chromatin domain data that was built from readDomBed function.
seqInfoChr	A Seqinfo object of the corresponding genome.
analysisMode	A character vector analysisMode=c("EnrichmentTest", "MLR", "MLRLasso", "MLRInter", "MLRInterLasso"). "EnrichmentTest" for enrichment test by simple logistic regression, "MLR" for multiple logistic regression, "MLRLasso" for multiple logistic regression using lasso parameter estimation, and "MLRInter" for multiple logistic regression with interaction terms provided by the interactionTerms parameter, and "MLRInter" for multiple logistic regression with interaction terms using lasso parameter estimation.
binSize	The size of the genomic bins (in bp) used for logistic regression computation. By default, binSize = 50.

<code>borderSize</code>	The size $\times 2$ of the window (in bp) surrounding the border between two domains. By default, <code>borderSize = 1000</code> .
<code>LRT</code>	If <code>TRUE</code> , a likelihood ratio test is used to assess the significance of beta parameters. By default, <code>LRT = FALSE</code> , and Wald's test is computed instead.
<code>interactionTerms</code>	A character vector <code>InteractionTerms</code> that contains the interaction terms between the genomic features. For instance, if there are two genomic features "F1" and "F2", the user can add the interaction term "F1:F2" in the same manner as in an <code>lm</code> or <code>glm</code> object.
<code>verbose</code>	If <code>verbose = TRUE</code> , then each step of the method is printed. By default, <code>verbose = FALSE</code> .

Value

The object contains 8 attributes:

<code>Enrich</code>	Enrichment test (simple logistic regression) result for each genomic feature.
<code>MLR</code>	Multiple logistic regression results for all genomic features.
<code>MLRLasso</code>	Multiple logistic regression results for all genomic features. Parameters of the logistic regression model are learned by lasso.
<code>MLRInter</code>	Multiple logistic regression results for all genomic features together with interaction terms that are provided in the character vector <code>interactionTerms</code> .
<code>MLRInterLasso</code>	Multiple logistic regression results for all genomic features together with interaction terms that are provided in the character vector <code>interactionTerms</code> . Parameters of the logistic regression model are learned by lasso.
<code>Mat</code>	Matrix used for all logistic regression computations.
<code>MultGLM</code>	The <code>glm</code> object from the multiple logistic regression model without interaction terms.
<code>InterGLM</code>	The <code>glm</code> object from the multiple logistic regression model with interaction terms.
<code>CorGF</code>	Correlation matrix between genomic features.

Author(s)

Raphael Mourad

Examples

```
# Load library HiCfeat
library(HiCfeat)

# Load data example.
data(dataExample)

# Multiple logistic regression estimated by maximum likelihood
print("Multiple logistic regression estimated by maximum likelihood")
BA_res=borderAnalysisFun(genomicFeatureList.GR=dataExample$GenomicFeatureList.GR,
```

```
GFDataType="bed", annotNames=dataExample$AnnotNames, domains.GR=dataExample$Domains.GR,  
seqInfoChr=dataExample$SeqInfoChr, analysisMode="MLR", binSize=500, borderSize=5e3,  
LRT=FALSE, interactionTerms="", verbose=FALSE)
```

dataExample

Read chromatin domain bed file

Description

This dataExample object comprises three objects: GenomicFeatureList.GR, AnnotNames and Domains.GR. GenomicFeatureList.GR is a list of GRanges objects, one for each genomic feature to analyze. AnnotNames is a vector of the genomic feature names. Domains.GR is a GRanges object for topologically associating domains.

Topologically associating domains (object Domains.GR) were computed from Hi-C experiments on *Drosophila melanogaster* Kc167 cells from Gene Expression Omnibus (GEO) accession GSE62904. ChIP-seq data (object GenomicFeatureList.GR) on *Drosophila melanogaster* Kc167 cells were downloaded from Gene Expression Omnibus (GEO) accessions GSE30740, GSE42085 and GSE54529. Only data for chromosome 2L are provided.

Usage

```
data(dataExample)
```

Author(s)

Raphael Mourad

References

Kevin Van Bortle, Michael H. Nichols, Li Li, Chin-Tong Ong, Naomi Takenaka, Zhaohui S. Qin, and Victor G. Corces. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biology*, 15(5):R82+, June 2014.

Li Li, Xiaowen Lyu, Chunhui Hou, Naomi Takenaka, Huy Q. Nguyen, Chin-Tong Ong, Caelin Cubenas-Potts, Ming Hu, Elissa P. Lei, Giovanni Bosco, Zhaohui S. Qin and Victor G. Corces. Widespread rearrangement of 3D chromatin organization underlies Polycomb-mediated stress-induced silencing. *Molecular Cell*, 15:S1097-2765, March 2015.

readDomBed	<i>Read chromatin domain bed file</i>
------------	---------------------------------------

Description

Function to import a chromatin domain bed file. The bed file must contain coordinates of the chromatin domains. It can be for instance topologically associating domain (TAD) coordinates. In the bed file, the first column is the chromosome, the second column is the starting position of the domain, the third column is the ending position of the domain. The fourth column (optional) is the annotation of the domain (for instance: active, PcG, ...).

Usage

```
readDomBed(domainBedFile, seqInfoChr)
```

Arguments

domainBedFile	The name of the bed file. If it does not contain an absolute path, the file name is relative to the current working directory.
seqInfoChr	A Seqinfo object for the corresponding genome.

Value

A GRanges object is returned.

Author(s)

Raphael Mourad

Examples

```
dom_file <- system.file("inst/extdata", "TAD_Corces_1kb_type_p1.bed", package="HiCfeat")
data(dataExample)
seqInfoChr=dataExample$SeqInfoChr
domBed=readDomBed(dom_file, seqInfoChr)
```

readGFBed	<i>Read genomic feature bed file</i>
-----------	--------------------------------------

Description

Function to import a genomic feature bed file. The bed file must contain coordinates of the genomic features. It can be for instance ChIP-seq peak coordinates or functional element coordinates.

Usage

```
readGFBed(GFBedFile, seqInfoChr)
```

Arguments

- GFBedFile The name of the bed file. If it does not contain an absolute path, the file name is relative to the current working directory.
- seqInfoChr A Seqinfo object for the corresponding genome.

Value

A GRanges object is returned.

Author(s)

Raphael Mourad

Examples

```
bed_file <- system.file("inst/extdata", "BEAF32_Kc167_CS_dm3.bed", package="HiCfeat")
data(dataExample)
seqInfoChr=dataExample$SeqInfoChr
GFBed=readGFBed(bed_file, seqInfoChr)
```

readGFWig *Read genomic feature bigwig file*

Description

Function to import a genomic feature wig or bigWig file. The file can contain for instance normalized ChIP-seq values ($\log_2(\text{ChIP}/\text{Input})$ values).

Usage

```
readGFWig(GFWigFile, seqInfoChr)
```

Arguments

- GFWigFile The name of the wig or bigWig file. If it does not contain an absolute path, the file name is relative to the current working directory.
- seqInfoChr A Seqinfo object for the corresponding genome.

Value

A GRanges object is returned.

Author(s)

Raphael Mourad

Examples

```
# Same use as for readGFBed(). An example wig or bigWig file is not provided in the package  
# because these files have large memory sizes.
```

Index

[averagePerBin](#), 2

[borderAnalysisFun](#), 3

[dataExample](#), 5

[HiCfeat \(HiCfeat-package\)](#), 2

[HiCfeat-package](#), 2

[readDomBed](#), 6

[readGFBed](#), 6

[readGFwig](#), 7