Package 'RHiCDB'

May 31, 2018

Type Package

Title RHiCDB detects contact domain boundaries(CDB) on Hi-C matrix

Description RHiCDB detects contact domain boundaries(CDB) on Hi-C matrix.

Version 1.0

Date 2018-04-27

Author Chen Fengling, Yuan bo et al.

Maintainer Chen Fengling <cfl15@mails.tsinghua.edu.cn>

License GPL (>=2)

Depends R (>= 3.2)

Imports pracma,limma,Matrix,gridExtra,rasterVis,lattice

RoxygenNote 6.0.1

R topics documented:

RHiCDB 1
visHiCDB 3

Index **5**

`RHiCDB` *Detect CDBs and differential CDBs on Hi-C heatmap.*

Description

HiCDB using Hi-C contact matrix to detect Hi-C contact domain boundaries(CDBs).It outputs annotated CDBs, differential CDBs on the chosen options

Usage

```
RHiCDB(hicfile, resolution, chrsizes, ref = "no", outdir, mind, wd, wdsize)
```

Arguments

hicfile	hicfile is the directory of the intra-chromosome Hi-C matrixes with sparse or dense format. The intra-chromosome matrix must be named as "chr+number.matrix" according to the chromosome order like 'chr1.matrix','chr2.matrix',..., 'chr23.matrix'. As HiCDB matches "chr*.matrix" to recognize the Hi-C matrix, avoid to use the "chr*.matrix" as the name of other files. The intra-chromosome matrix could be in a dense (a NxN matrix) or sparse (a Kx3 table,Rao et al.) format. If you want to detect CDB on one sample,set hicfile as 'SAMPLE_DIR'. If ref is not set, this function will output all the local maximum peaks. If ref is set, this function will output local maximum peaks and final CDBs. If you want to detect differential CDBs, ref is required to decide the cut off on CDB detection. If you don't have replicate, set hicfile as list('SAMPLE1','SAMPLE2'). This function will first perform CDB detection on each sample and then compare the difference between their final CDBs by intersection. If replicates is provided, set hicfile as list(c('SAMPLE1_rep1','SAMPLE1_rep2'),c('SAMPLE2_rep1','SAMPLE2_rep2')). The function will find CDBs on each sample with merged Hi-C matrix, calculate aRI score on each replicates, then decide a CDB as differetial or not by statistical test on aRI scores of each CDB. If ref is 'hg38' or 'hg19', CDBs will also be annotated as conserved or not conserved.
resolution	resolution of Hi-C matrix. This is required.
chrsizes	Ordered chromosome sizes of the genome. Optional setting is 'hg19', 'hg38', 'mm9', 'mm10' or any other chromosome size files which can be generated following the instructions in annotation/README.md. This is required.
ref	reference CTCF motif locs on the genome. If it is set, the output will use the GSEA-like methods to decide the cutoff. Default is 'no'. Choices are : 'no' 'hg19' 'hg38' 'mm9' 'mm10' or other customfile for example 'genome.txt' made from utility/motifanno.sh Example for 'genome.txt': #'chr motifcenter-locus 10 15100928 10 15188593
outdir	The output directory. Default will be the directory of the first sample.
mind	Minimum local maximum peak distance (measured by bin), or minimum separation between local maximum peaks, specified as a positive integer scalar. Use this argument to have findpeaks ignore small peaks that occur in the neighborhood of a larger peak.
wd	The smallest window sizes.
wdsize	The number of different window size. The whole window size scale will be wd:(wd+wdsize).Default will be 6.

Details

A. Possible outputs

1.CDB.txt

2.localmax.txt: all the local maximum peaks detected before cutoff decision. User can decide custum CDB cutoff upon this file.

3.EScurve.png: CTCF motif enrichment on ranked local maximum peaks.

4.aRI.txt: average RI score for each genomic bin.

5.LRI.txt: LRI score for each genomic bin.

B. default value for 'mind','wd' on different resolution

resolution mind wd wdsize

10k 4 3 6

40k 2 1 6

5k 8 6 8

C. HiCDB will perform a KR normlization if the data is raw counts.

Author(s)

Implemented by Fengling Chen

Any suggestions and remarks might be addressed to Fengling Chen:cfl15@mails.tsinghua.edu.cn

Examples

```
      1. Output all the local maximum peaks and let customers to decide the cutoff.
      RHiCDB('sample1/',10000,chrsizes='custom_chrsizes.txt');
      RHiCDB('sample1/',10000,chrsizes='custom_chrsizes.txt',outdir='sample1/outputs/');
      2. Use GSEA-like methods to decide the cutoff
      RHiCDB('sample1/',10000,chrsizes='hg19',ref='hg19');
      RHiCDB('sample1/',10000,chrsizes='custom_chrsizes.txt',ref='custom_motiflocs.txt');
      3. To detect differential CDBs
       RHiCDB(list('sample1','sample2'),10000,'hg19',ref='hg19');
       RHiCDB(list(c("sample1_rep1","sample1_rep2"),c("sample2_rep1","sample2_rep2")),
  + 10000,'hg19',ref='hg19');
```

`visHiCDB` *Visualization of CDBs or differential CDBs on Hi-C maps*

Description

visHiCDB uses Hi-C raw contact matrix and CDBs as input and outputs figures of CDBs or differential CDBs on Hi-C maps

Usage

```
visHiCDB(hicfile, CDBfile, resolution, chr, startloc, endloc, outdir)
```

Arguments

`hicfile`	hicfile is the file names of the intra-chromosome matrixes. The intra-chromosome matrix could be in a dense (a NxN matrix) or sparse (a Kx3 table,Rao et al.) format. Show CDBs on one sample,set hicfile as 'SAMPLE_File'. Show differebtial CDBs in two samples, set hicfile as list('SAMPLE1_FILE','SAMPLE2_FILE').
`CDBfile`	CDBfile should be is file name of the CDB files. Show CDBs on one sample,set CDBfile as 'SAMPLE_CDB'. Show differebtial CDBs in two samples, set CDBfile as list('SAMPLE1_CDB','SAMPLE2_CDB'). The CDB file should be formated as the output file of HiCDB.
`resolution`	resolution of Hi-C matrix. This is required.
`chr, startloc, endloc`	numeric observation locus on Hi-C map.This is required.
`outdir`	The output direction. Default will be the directory of the first sample.

Details

This function outputs a pdf figure showing CDBs on a Hi-C map on desired locus. Conserved CDBs are marked as dark blue.

Author(s)

Implemented by Fengling Chen

Any suggestions and remarks might be addressed to Fengling Chen:cfl15@mails.tsinghua.edu.cn

Examples

```
## Show CDB on single Hi-C map
visHiCDB('sample1/chr18.matrix','CDB1.txt',40000,18,25000000,31150000)
## Show differential CDBs
visHiCDB(list('sample1/chr18.matrix','sample2/chr18.matrix'),
   +list('CDB1.txt','CDB2.txt'),40000,18,25000000,31150000)
```

Index

RHiCDB, 1

visHiCDB, 3