rGMAP

September 6, 2018

Type Package

Title Call hierarchical chromatin domains from HiC matrix by GMAP

Version 1.3.1

Date 08-23-2018

Author Wenbao Yu

Maintainer Wenbao Yu <yuw1@email.chop.edu>

Description Call hierarchical chromatin domains from HiC contact map by Gaussian Mixture model And Proportion test

BugReports https://github.com/wbaopaul/rGMAP/issues

License GPL (>= 2)

LazyData TRUE

Imports data.table,
ggplot2,
mclust,
EMD,
caTools,
Matrix,
Rcpp (>= 0.12.5)

LinkingTo Rcpp

RoxygenNote 6.1.0

Suggests knitr,
rmarkdown

VignetteBuilder knitr

R topics documented:

data_simu *generate simulated hic_mat and true tads*

Description

generate simulated hic_mat and true tads

Usage

```
data_simu(stype = "poisson-dist", nratio = 2.5, mu0 = 200,
  resl = 1)
```

Arguments

stype	One of four types of simulated data in the manuscript: poission-dist, poission-dist-hier, nb-dist, nb-dist-hier; poission- or nb- indicates poission distribution or negative bionomial distribution -hier indicated subtads are generated nestly
nratio	The effect size between intra- and inter domain, larger means higher intra-tad contacts
mu0	The mean parameter, default 200
resl	Resolution, default set to 1

Value

A list includes following elements:

hic_mat	n by n contact matrix
hierTads	True heirarchical domains
tads_true	True TADs

hic_rao_IMR90_chr15 *Normalized Hi-C data for IMR90, chr15 with resolution 10kb.*

Description

Normalized Hi-C data for IMR90, chr15 with resolution 10kb.

Usage

```
hic_rao_IMR90_chr15
```

Format

A data table with 3 variables:

n1 bin 1

n2 bin 2

count normalized counts

Source

Rao et al., Cell 2014, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping

---

`plotdom` *visualize hierarchical domains*

Description

visualize hierarchical domains

Usage

```
plotdom(hic_dat, hiertads_gmap, start_bin, end_bin, cthr = 20,
  resl = 10000)
```

Arguments

`hic_dat`	hic contact matrix for a given chromosome, either a n by n matrix, or a 3 columns data.frame <bin1> <bin2> <counts>
`hiertads_gmap`	the hierarchical domains called by GMAP
`start_bin`	the start bin of the genome
`end_bin`	the end bin of the genome
`cthr`	the upper bound count threshold for color, default 20
`resl`	reslution of Hi-C data, default 10000

`rGMAP` *Detect hierarchical choromotin domains by GMAP*

Description

Detect hierarchical choromotin domains by GMAP

Usage

```
rGMAP(hic_mat, index_file = NULL, resl = 10 * 10^3, logt = T,
  dom_order = 2, maxDistInBin = min(200, 2 * 10^6/resl), min_d = 25,
  max_d = 100, min_dp = 5, max_dp = 10, hthr = 0.95, t1thr = 0.5)
```

Arguments

hic_mat	• For single chromosome, supports three types of format: – a 3-column Hi-C contact matrix, with columns the i_th, j_th bin of a chromosom and the corresponding contact number – a n by n matrix, with <i,j>th element corresponding to contact number between the i_th and j_th bin of the chromosome – a text file of the above two types of data • For multiple chromosomes, a index_file indicates genomic coordinate for each hic bin should be provided
index_file	A 4-columns tab/space delimited text file indicates the genomic coordinates for each bin (compatible with HiC-Pro); with columns *bin_chr bin_start bin_end bin_id*
resl	The resolution (bin size), default 10kb
logt	Do log-transformation or not, default TRUE
dom_order	Maximum level of hierarchical structures, default 2 (call TADs and subTADs)
maxDistInBin	Only consider contact whose distance is not greater than maxDistInBIn bins, default 200 bins (or 2Mb)
min_d	The minimum d (d: window size), default 25
max_d	The maximum d (d: window size), default 100
min_dp	The minmum dp (dp: lower bound of tad size), defalt 5
max_dp	The maximum dp (dp: lower bound of tad size), defalt 10. min_d, max_d, min_dp and max_dp should be specified in number of bins
hthr	The lower bound cutoff for posterior probability, default 0.95
t1thr	Lower bound for t1 for calling TAD, default 0.5 quantile of test statistics of TADs, 0.9 of subTADs

Value

A list includes following elements:

tads	A data frame with columns start, end indicates the start and end coordinates of each domain, respectively
hierTads	A data frame with columns start, end, dom_order, where dom_order indicates the hierarchical status of a domain, 1 indicates tads, 2 indicates subtads, and so on
params	A data frame gives the final parameters for calling TADs

Examples

```
## On simulated data:
library(rGMAP)
simu_res = data_simu('poisson-dist-hier')
true_domains = simu_res$hierTads
simu_mat = simu_res$hic_mat
predicted_domains = rGMAP(simu_mat, resl = 1)$hierTads
true_domains
predicted_domains

## On an real data example
```

```
hic_rao_IMR90_chr15   # normalized Hi-C data for IMR90, chr15 with resolution 10kb
res = rGMAP(hic_rao_IMR90_chr15, resl = 10 * 1000, dom_order = 2)
names(res)
#quickly visualize some hierarchical domains
pp = plotdom(hic_rao_IMR90_chr15, res$hierTads, 6000, 7000, 30, 10)
pp$p2
```

Index