Package ‘rcpphmmparclip’

October 15, 2013

Type Package

Title Rcpp functions for analysing PAR-CLIP using the HMM

Version 0.1.0

Date 2013-10-15

Author Jonghyun Yun

Maintainer Jonghyun Yun <jonghyun.yun@utsouthwestern.edu>

Description This package provides essential functions for the forward-backward Gibbs sampler to analyse PAR-CLIP data using the HMM.

Depends Rcpp (>= 0.9.3)

License GPL (>= 2.1)

R topics documented:

rcpphmmparclip-package ................................................. 1
bases ........................................................................... 2
fbgibbs ....................................................................... 3

Index

rcpphmmparclip-package

Rcpp functions for analysing PAR-CLIP using the HMM

Description

This package provides essential functions for the forward-backward Gibbs sampler to analyse PAR-CLIP data using the HMM.
Details

To install this package, windows users may require to install Rtools (http://cran.r-project.org/bin/windows/Rtools/), which will help to compile the Rcpp functions in your machine. It is recommended to install both R and Rtools in directories whose names have no space. More details about how to build the package is provided in Rcpp-FAQ (http://dirk.eddelbuettel.com/code/rcpp/Rcpp-FAQ.pdf).

To install the package,

install.packages('rcpphmmparclip_0.1.0.tar.gz',type='source')

Author(s)

Jonghyun Yun <jonghyun yun@utsouthwestern.edu>

References


See Also

bases, fbgibbs

Examples

```r
require(rcpphmmparclip)
require(doMC)
require(parallel)

tcore = detectCores()
cl <- makeCluster( tcore )
registerDoMC()

data(bases)

mc = fbgibbs(bases,tot=100,nsv=tcore);

# print sites with the posterior probability >= postcut.
postcut = 0.8;
ranking(bases,mc,postcut);
# estimate FDR on these sites.
fdr(mc,postcut);

stopCluster(cl)
rm(cl)
```

bases An example PAR-CLIP data set

Description

This is a short example of PAR-CLIP dataset. Aligned read sequences are converted to read and mutation counts in each genomic location. Only T to C substitutions are kept, and non-clustered reads are discarded.
Usage

data(bases)

Format

tag integer Read counts
mutant integer Mutation counts (T to C substitutions)
region_id string ID that is assigned to a CLIP cluster
chr string Chromosome
strand string Strand information
nt string Nucleotide sequence
pos integer Genomic location

Examples

require(rcpphmmparclip)
require(ggplot2)
data(bases)
attach(bases)
rid = 85974;
ridx = (region_id == rid);
rlen = sum(ridx);
x = rep(as.factor(pos[ridx]),2);
y = c(tag[ridx] - mutant[ridx],mutant[ridx]);
factor = as.factor(c(rep("read",rlen),rep("mutation",rlen)))
qplot(x,geom="bar", weight=y, fill = factor, xlab='Genomic Location',
title='Graphical representation of a CLIP cluster')

fbgibbs

The forward-backward Gibbs sampler

Description

This function estimates posterior probabilities of binding sites, posterior means of parameters. The function also provides the posterior predictive checking.

Usage

fbgibbs(data, c = 3, p = c(0.04, 0.02, 0.3), pi = c(0.5, 0.5, 0),
        fkab = t(matrix(c(1,-1,0,
                          1,1,0.05,
                          1,1,0.05),3,3)),
        K_T = matrix(c(0.98, 0.02, 0.00,
                       0.05, 0.92, 0.03,
                       0.00, 0.99, 0.01),3,3),
        K_N2 = c(0.05, 0.95, 0.00),
        phi_T = c(0.90, 0.1, 0.00),
        epsilon = 0.195, delta = 0.005, tot = 10000, burnp = 0.5,
        nsv = 1, doppc = FALSE, doprintout = FALSE)
Arguments

data PAR-CLIP data to be analysed. See ...
c The truncation value for read counts.
p,\pi Initial values of mutation probabilities \( p_s \) and zero-inflated probabilities \( \pi_s \).
fkab Initial values of parameters in the beta geometric distribution.
K_T Initial values of transition matrix \( K_T \) on T-sites.
K_N Initial values of the 2nd row of the transition matrix \( K_N \) on non T-sites.
phi_T Initial values of the initial distribution on T-sites of hidden Markov chains.
epsil The upper bound \( \epsilon \) of mutation probabilities on non-binding sites.
delta The lower bound \( \epsilon + \delta \) of mutation probabilities on binding sites.
tot The maximum number of MCMC iterations.
burnp The proportion of MCMC chains to be burnt out.
nsv The number of cores to be used for parallel computing.
doppc If TRUE, the posterior predictive checking is carried out. However, this option will slow down the FB Gibbs sampler.
doprintout If TRUE, summaries for each MCMC chains will be provided.

Value

IP [.3] Posterior probability of the binding site
ppc The p-value for the posterior predictive checking
sumc posterior means of parameters
sumc$bg posterior means of parameters in the BG. [.2]: \( \mu_s \) and [.3]: \( \eta_s \)

See Also

bases

Examples

```r
require(rcpphmmparclip)
require(doMC)
require(parallel)

tcore = detectCores()
c1 <- makeCluster( tcore )
registerDoMC()

data(bases)

mc = fbgibbs(bases,tot=100,nsv=tcore);

# print sites with the posterior probability >= postcut.
postcut = 0.8;
ranking(bases,mc,postcut);
# estimate FDR on these sites.
fdr(mc,postcut);

stopCluster(cl)
rm(cl)
```
Index

bases, 2, 2, 4
fbgibbs, 2, 3
rcpphmmparclip-package, 1