Package ‘iChip’

April 25, 2017

Title Bayesian Modeling of ChIP-chip Data Through Hidden Ising Models
Version 1.30.0
Date 2012-9-26
Depends R (>= 2.10.0)
Imports limma
Author Qianxing Mo
Maintainer Qianxing Mo <qmo@bcm.edu>
Description This package uses hidden Ising models to identify enriched genomic regions in ChIP-chip data. It can be used to analyze the data from multiple platforms (e.g., Affymetrix, Agilent, and NimbleGen), and the data with single to multiple replicates.
LazyData yes
License GPL (>= 2)
bioCViews ChIPchip, OneChannel, AgilentChip, Microarray
NeedsCompilation yes

R topics documented:
enrichreg .................................................. 1
iChip1 .................................................. 3
iChip2 .................................................. 5
lmtstat .................................................. 7
oct4 .................................................. 9
p53 .................................................. 9

Description

A function used to call and merge enriched probes to enriched regions using the posterior probability calculated by iChip2 or iChip1 functions at certain posterior probability and false discovery rate (FDR) cutoffs.
Usage

enrichreg(pos,enrich,pp,cutoff,method=c("ppcut","fdrcut"),maxgap=500)

Arguments

pos A n by 2 matrix or data frame. Rows correspond to probes. The first column of the matrix contains chromosome IDs; the second column contains the genomic positions.
enrich A vector containing the probe enrichment measurements.
pp A vector containing the posterior probabilities returned by iChip2 or iChip1.
cutoff The cutoff value (a scalar) used to call enriched probes. If use posterior probability as a criterion (method="ppcut"), a probe is said to be enriched if its pp is greater than the cutoff. If use FDR as a criterion (method="fdrcut"), probes are said to be enriched if the probe-based FDR is less than the cutoff. The FDR is calculated using a direct posterior probability approach (Newton et al., 2004).
method 'ppcut' or 'fdrcut'.
maxgap The criterion used to merge enriched probes. If the genomic distance of adjacent probes is less than maxgap, the probes will be merged into the same enriched regions.

Value

A data frame with rows corresponding to enriched regions and columns corresponding to the following:

chr Chromosome IDs. For human genome, 23 and 24 denote X and Y, respectively.
gstart The start genomic position of the enriched region.
gend The end genomic position of the enriched region.
rrstart The row number for gstart in the position matrix.
rend The row number for gend in the position matrix.
peakpos The peak genomic position of the enriched region where the probe has the largest enrichment value.
meanpp The mean posterior probability of the probes in the enriched region.
maxpp The maximum posterior probability of the probes in the enriched region.
nprobe The number of probes in the enriched regions. nprobe = rend - rstart + 1

Author(s)

Qianxing Mo <qmo@bcm.edu>

References


See Also

iChip2, iChip1, lmtstat

Examples

library(iChip)
library(limma)

#Analyze the p53 data (average resolution is about 35 bps)
#uncommenting the following code for running

#data(p53)
#p53lmt = lmtstat(p53[,9:14],p53[,3:8])
#p53Y = cbind(p53[,1],p53lmt)
#p53res=iChip2(Y=p53Y,burnin=2000,sampling=10000,winsize=2,sdcut=2,beta=2.5)
#enrichreg(pos=p53[,1:2],enrich=p53lmt,pp=p53res$pp,cutoff=0.9,
# method="ppcut",maxgap=500)
#enrichreg(pos=p53[,1:2],enrich=p53lmt,pp=p53res$pp,cutoff=0.01,
# method="fdrcut",maxgap=500)

iChip1

Bayesian modeling of ChIP-chip data through hidden Ising models

Description

Function iChip1 implements the algorithm of modeling ChIP-chip data through a standard hidden Ising model.

Usage

iChip1(enrich,burnin=2000,sampling=10000,sdcut=2,beta0=3,
       minbeta=0,maxbeta=10,normsd=0.1,verbose=FALSE)

Arguments

enrich A vector containing the probe enrichment measurements. The measurements must be sorted, firstly by chromosome and then by genomic position. The measurements could be log2 ratios of the intensities of IP-enriched and control samples for a single replicate, or summary statistics such as t-like statistics or mean differences for multiple replicates. We suggest to use the empirical Bayesian t-statistics implemented in the limma package for multiple replicates. Note, binding probes must have a larger mean value than non-binding probes.

burnin The number of MCMC burn-in iterations.

sampling The number of MCMC sampling iterations. The posterior probability of binding and non-binding state is calculated based on the samples generated in the sampling period.

sdcut A value used to set the initial state for each probe. The enrichment measurements of a enriched probe is typically several standard deviations higher than the global mean enrichment measurements.
The initial parameter used to control the strength of interaction between probes, which must be a positive value. A larger value of beta represents a stronger interaction between probes. The value for beta0 could not be too small (e.g. < 1.0). Otherwise, the Ising system may not be able to reach a super-paramagnetic state.

minbeta

The minimum value of beta allowed.

maxbeta

The maximum value of beta allowed.

normsd

iChip1 uses a Metropolis random walk proposal for sampling from the posterior distributions of the model parameters. The proposal distribution is a normal distribution with mean 0 and standard deviation specified by normsd.

verbose

A logical variable. If TRUE, the number of completed MCMC iterations is reported.

Value

A list with the following elements.

pp

The posterior probabilities of probes in the binding/enriched state. There is a strong evidence to be a binding/enriched probe if the probe has a posterior probability close to 1.

beta

The posterior samples of the interaction parameter of the Ising model.

mu0

The posterior samples of the mean measurement of the probes in the non-binding/non-enriched state.

mu1

The posterior samples of the mean measurement of the probes in the binding/enriched state.

lambda

The posterior samples of the precision of the enrichment measurements of the probes.

Author(s)

Qianxing Mo <qmo@bcm.edu>

References


See Also

iChip2, enrichreg, lmtstat

Examples

```r
# oct4 and p53 data are log2 transformed and quantile-normalized intensities
# Analyze the Oct4 data (average resolution is about 280 bps)
data(oct4)

### sort oct4 data, first by chromosome then by genomic position
oct4 = oct4[order(oct4[,1],oct4[,2]),]
```
# calculate the enrichment measurements --- the limma t-statistics
oct4lmt = lmtstat(oct4[,5:6],oct4[,3:4])

# Apply the standard Ising model to the ChIP-chip data
oct4res = iChip1(enrich=oct4lmt,burnin=1000,sampling=5000,sdcut=2,
beta0=3,minbeta=0,maxbeta=10,normsd=0.1)

# check the enriched regions detected by the Ising model using # posterior probability (pp) cutoff at 0.9 or FDR cutoff at 0.01
enrichreg(pos=oct4[,1:2],enrich=oct4lmt,pp=oct4res$pp,cutoff=0.9,
method="ppcut",maxgap=500)
enrichreg(pos=oct4[,1:2],enrich=oct4lmt,pp=oct4res$pp,cutoff=0.01,
method="fdrcut",maxgap=500)

# Analyze the p53 data (average resolution is about 35 bps)
# uncommenting the following code for running
# data(p53)
# must sort the data first
# p53 = p53[order(p53[,1],p53[,2]),]
# p53lmt = lmtstat(p53[,9:14],p53[,3:8])
# p53res = iChip1(p53lmt,burnin=1000,sampling=5000,sdcut=2,beta0=3,
# minbeta=0,maxbeta=10,normsd=0.1)
# enrichreg(pos=p53[,1:2],enrich=p53lmt,pp=p53res$pp,cutoff=0.9,
# method="ppcut",maxgap=500)
# enrichreg(pos=p53[,1:2],enrich=p53lmt,pp=p53res$pp,cutoff=0.01,
# method="fdrcut",maxgap=500)

---

**iChip2**  
_Bayesian modeling of ChIP-chip data through hidden Ising models_

**Description**

Function `iChip2` implements the method of modeling ChIP-chip data through a high-order hidden Ising model.

**Usage**

```r
iChip2(Y,burnin=2000,sampling=10000,winsize=2,sdcut=2,beta=2.5,verbose=FALSE)
```

**Arguments**

- **Y**: A n by 2 matrix or data frame. The first column of Y contains the chromosome IDs; the second column of Y contains the probe enrichment measurements. Y must be sorted, firstly by chromosome and then by genomic position. The probe enrichment measurements could be log2 ratios of the intensities of IP-enriched and control samples for a single replicate, or summary statistics such as t-like statistics or mean differences for multiple replicates. We suggest to use the
empirical Bayesian t-statistics implemented in the limma package for multiple replicates. Note, binding probes must have a larger mean value than non-binding probes.

**burnin**
The number of MCMC burn-in iterations.

**sampling**
The number of MCMC sampling iterations. The posterior probability of binding and non-binding state is calculated based on the samples generated in the sampling period.

**winsize**
The parameter to control the order of interactions between probes. For example, winsize = 2, means that probe \(i\) interacts with probes \(i-2, i-1, i+1\) and \(i+2\). A balance between high sensitivity and low FDR could be achieved by setting winsize = 2.

**sdcut**
A value used to set the initial state for each probe. The enrichment measurements of a enriched probe is typically several standard deviations higher than the global mean enrichment measurements.

**beta**
The parameter used to control the strength of interaction between probes, which must be a positive value. A larger value of beta represents a stronger interaction between probes. In general, high resolution array such as Affymetrix tiling arrays have relatively stronger probe interactions than low resolution array such as Agilent tiling arrays. For the second order Ising model (winsize = 2), the critical value of beta is around 1.0. For low resolution array data (e.g. 280 bp resolution), beta could be set to close to the critical value; For high resolution array data (e.g. 35 bp resolution), beta could be set to a value between 2 to 4. In general, choosing a large value of beta amounts to using a more stringent criterion for detecting enriched regions in ChIP-chip experiments.

**verbose**
A logical variable. If TRUE, the number of completed MCMC iterations is reported.

**Value**
A list with the following elements.

**pp**
The posterior probabilities of probes in the binding/enriched state. There is a strong evidence to be a binding/enriched probe if the probe has a posterior probability close to 1.

**mu0**
The posterior samples of the mean measurement of the probes in the non-binding/non-enriched state.

**mu1**
The posterior samples of the mean measurement of the probes in the binding/enriched state.

**lambda0**
The posterior samples of the precision of the enrichment measurements of the probes in the non-binding/non-enriched state.

**lambda1**
The posterior samples of the precision of the enrichment measurements of the probes in the binding/enriched state.

**Author(s)**
Qianxing Mo <qmo@bcm.edu>

**References**
See Also

iChip1, enrichreg, lmtstat

Examples

# oct4 and p53 data are log2 transformed and quantile-normalized intensities
#
# Analyze the Oct4 data (average resolution is about 280 bps)
data(oct4)

### sort oct4 data, first by chromosome then by genomic position
oct4 = oct4[order(oct4[,1],oct4[,2]),]
#
# calculate the enrichment measurements --- the limma t-statistics
oct41mt = lmtstat(oct4[,5:6],oct4[,3:4])
#
# prepare the data used for the Ising model
oct4Y = cbind(oct4[,1],oct41mt)
#
# Apply the second-order Ising model to the ChIP-chip data
oct4res=iChip2(Y=oct4Y,burnin=1000,sampling=5000,winsize=2,sdcut=2,beta=1.25)
#
# check the enriched regions detected by the Ising model using
# posterior probability (pp) cutoff at 0.9 or FDR cutoff at 0.01
enrichreg(pos=oct4[,1:2],enrich=oct41mt,pp=oct4res$pp,cutoff=0.9,
method="ppcut",maxgap=500)
enrichreg(pos=oct4[,1:2],enrich=oct41mt,pp=oct4res$pp,cutoff=0.01,
method="fdrcut",maxgap=500)

# Analyze the p53 data (average resolution is about 35 bps)
# uncommenting the following code for running
#
data(p53)
# must sort the data first
# p53 = p53[order(p53[,1],p53[,2]),]
# p531mt = lmtstat(p53[,9:14],p53[,3:8])
# p53Y = cbind(p53[,1],p531mt)
# p53res=iChip2(Y=p53Y,burnin=1000,sampling=5000,winsize=2,sdcut=2,beta=2.5)
#
# enrichreg(pos=p53[,1:2],enrich=p531mt,pp=p53res$pp,cutoff=0.9,
# method="ppcut",maxgap=500)
# enrichreg(pos=p53[,1:2],enrich=p531mt,pp=p53res$pp,cutoff=0.01,
# method="fdrcut",maxgap=500)
Description
A wrapper function used to calculate the empirical Bayes t-statistics (limma t-statistics) using functions in the limma package.

Usage
lmtstat(IP, CON)

Arguments
IP  Data matrix for IP-enriched samples, where the rows and columns correspond to the probes and sample replicates, respectively. The number of replicates must be greater than one. If CON is missing, IP is assumed to be in log-ratio format (e.g. log2(IP-enriched/control)). In this case, paired t-statistics are calculated. If CON is NOT missing, IP and CON are assumed to be the normalized intensities for the IP-enriched and control samples, respectively. In this case, two-sample t-statistics are calculated.

CON  Data matrix for control samples, where the rows and columns correspond to the probes and sample replicates, respectively. The number of replicates must be greater than one.

Value
Empirical Bayes t-statistics calculated using functions in the limma package.

Author(s)
Qianxing Mo <qmo@bcm.edu>

References

See Also
enrichreg, iChip2, iChip1

Examples
library(limma)
# load the log2 transformed and quantile-normalized Oct4 data
data(oct4)
oct4[1:3,]
# calculate the enrichment measurements --- two-sample limma t-statistics
oct4lmt1 = lmtstat(oct4[,5:6],oct4[,3:4])
# calculate paired limma t-statistics for the data that are in
# the log-ratio format (e.g., log2(IP-enriched/control))
oct4

\[ \text{oct4log2r} = \text{oct4[,5:6]} - \text{oct4[,3:4]} \]
\[ \text{oct4lmt2} = \text{lmtstat(oct4log2r)} \]

<table>
<thead>
<tr>
<th>oct4</th>
<th>Oct4 data</th>
</tr>
</thead>
</table>

**Description**

This is a subset of the Oct4 data containing 12584 probes on chromosome 20. The data were log2 transformed and quantile-normalized.

**Usage**

data(oct4)

**Source**

http://jura.wi.mit.edu/young\_public/hESregulation/Data\_download.html

**References**


<table>
<thead>
<tr>
<th>p53</th>
<th>p53 data</th>
</tr>
</thead>
</table>

**Description**

This is a subset of the p53 data containing 10000 probes on chromosome 22. The data were log2 transformed and quantile-normalized.

**Usage**

data(p53)

**Source**

http://www.gingeras.org/affy\_archive\_data/publication/tfbs/

**References**

# Index

*Topic **datasets**
  - oct4, 9
  - p53, 9

*Topic **models**
  - enrichreg, 1
  - iChip1, 3
  - iChip2, 5
  - lmtstat, 7
  - enrichreg, 1, 4, 7, 8
  - iChip1, 3, 3, 7, 8
  - iChip2, 3, 4, 5, 8
  - lmtstat, 3, 4, 7, 7
  - oct4, 9
  - p53, 9