Package 'enRich'

March 5, 2020

Type Package

Title Analysis of Multiple ChIP-Seq Data

Version 3.1

Date 2020-02-03

Author Yanchun Bao, Veronica Vinciotti

Maintainer Yanchun Bao <ybaoa@essex.ac.uk>

Description Joint statistical modelling of ChIP-seq data, accounting for technical/biological replicates, multiple conditions and different ChIP efficiencies of the individual experiments. <DOI:10.1186/1471-2105-14-169>. <DOI:10.1093/biostatistics/kxt047>.

Depends R(>= 3.2.2), parallel

License GPL-2

LazyLoad yes

RoxygenNote 7.0.2

Encoding UTF-8

NeedsCompilation yes

Repository CRAN

Date/Publication 2020-03-05 12:20:05 UTC

R topics documented:

enRich-package *Analysis of multiple ChIP-seq data.*

Description

enRich is an R package that performs a joint statistical modelling of ChIP-seq data, accounting for technical/biological replicates, multiple conditions and the different IP efficiencies of individual experiments.

Details

Package:	enRich
Type:	Package
Version:	3.1
Date:	2020-02-03
Depends:	R(>= 3.2.2), parallel
License:	GPL (>=2)
LazyLoad:	yes

mainfunctions: mix, mix.joint, mrf, mrf.joint, enrich.mix, enrich.mrf

Author(s)

Yanchun Bao <ybaoa@essex.ac.uk> and
Veronica Vinciotti <veronica.vinciotti@brunel.ac.uk>

Maintainer: Yanchun Bao <ybaoa@essex.ac.uk>

References

Bao et al. Accounting for immunoprecipitation efficiencies in the statistical analysis of ChIP-seq data. BMC Bioinformatics 2013, 14:169 DOI:10.1186/1471-2105-14-169.
Bao et al. Joint modelling of ChIP-seq data via a Markov random field model, Biostatistics 2014, 15(2):296-310 DOI:10.1093/biostatistics/kxt047.

enrich.mix *Detection of enriched and differentially bound regions for fitting results of* mix *and* mix.joint.

Description

enrich.mix returns the enriched regions or differentially bound regions using the mix or the mix.joint model, by controlling a given FDR level. enrich.mix also calculates the IP efficiencies for each experiment.

Usage

```
enrich.mix(object, analysis = "joint", differential = FALSE,
    diff.vec = NULL, cr = 0.05, crdiff = 0.05)
```

Arguments

object	The output of mix if analysis="separate" or of mix.joint if analysis="joint".
analysis	A character variable. Default value is "joint" and the object should be the output of mix.joint. If analysis="separate", then the object should be the output of mix.
differential	A logical variable. If TRUE, the function will compute the posterior probability of differential binding of any two experiments or two conditions, as specified by diff.vec. Default value is FALSE.
diff.vec	A numeric vector. If differential = TRUE, diff.vec must be given to show which experiments are to be used in the comparison. At the moment, this is restricted to two conditions (e.g. two proteins at the same time point), so the value for diff.vec should be only 0, 1, 2, where 0 indicates which experiments are not to be used in the analysis, 1 and 2 stand for conditions 1 and 2, respectively. diff.vec should be of the same length as the number of experiments in object.
cr	A numeric variable. The level of FDR for identifying the enriched regions.
crdiff	A numeric variable. The level of FDR for identifying the differentially bound regions.

Value

enrich	The list of enriched regions for each condition at the chosen FDR. Note that there is only one list of enriched regions for replicates, if a joint model is used.
diffenrich1	The list of regions bound only by condition 1.
diffenrich2	The list of regions bound only by condition 2.
ppx1	A n x p matrix of posterior probabilities of enrichment for each region and each condition. ppx0=1-ppx1.
X	A n x p matrix of enrichment for each region and each condition, at the given FDR cutoff (1: enriched, 0: not-enriched).
diffprob1	A n-dimensional vector of posterior probabilities of differential binding for the two conditions under study; diffprob0=1-diffprob1.
diffX1	A n-dimensional index of regions bound only by condition 1 (0: not bound, 1: bound).
diffX2	A n-dimensional index of regions bound only by condition 2.
IPE	A p-dimensional vector of estimated IP efficiency values for each experiment.

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Accounting for immunoprecipitation efficiencies in the statistical analysis of ChIP-seq data. BMC Bioinformatics 2013, 14:169 DOI:10.1186/1471-2105-14-169.

See Also

See also mix, mix.joint

Examples

```
tempdir()
data(p300cbp.1000bp)
exp.label=c("CBPT0", "CBPT301", "CBPT302", "p300T0",
    "p300T301", "p300T302", "WangCBP", "Wangp300")

## Simple examples -- only two experiments and first 5000 observations
CBPT30=list()
CBPT30$region=p300cbp.1000bp$region[1:5000,]
CBPT30$count=p300cbp.1000bp$count[1:5000,2:3]
Poisfit.simple<-mix(CBPT30, method="Poisson", exp.label=exp.label[c(2,3)])
enrich.mix.simple<-enrich.mix(Poisfit.simple, analysis="separate")
```

---

enrich.mrf *Detection of enriched and differentially bound regions for fitting results of* mrf *and* mrf.joint.

Description

enrich.mrf returns the enriched regions or differentially bound regions using the mrf or the mrf.joint model, by controlling a given FDR level. enrich.mrf also calculates the IP efficiencies for each experiment.

Usage

```
enrich.mrf(object, analysis = "joint", differential = FALSE,
    diff.vec = NULL, cr = 0.05, crdiff = 0.05)
```

Arguments

object	The output of mrf if analysis="separate" or of mrf.joint if analysis="joint".
analysis	A character variable. Default value is "joint" and the object should be the output of mrf.joint. If analysis="separate", then the object should be the output of mrf.
differential	A logical variable. If TRUE, the function will compute the posterior probability of differential binding of any two experiments or two conditions, as specified by diff.vec. Default value is FALSE.

diff.vec	A numeric vector. If differential = TRUE, diff.vec must be given to show which experiments are to be used in the comparison. At the moment, this is restricted to two conditions (e.g. two proteins at the same time point), so the value for diff.vec should be only 0, 1, 2, where 0 indicates which experiments are not to be used in the analysis, 1 and 2 stand for conditions 1 and 2, respectively. diff.vec should be of the same length as the number of experiments in object.
cr	A numeric variable. The level of FDR for identifying the enriched regions.
crdiff	A numeric variable. The level of FDR for identifying the differentially bound regions.

Value

enrich	The list of enriched regions for each condition at the chosen FDR. Note that there is only one list of enriched regions for replicates, if a joint model is used.
diffenrich1	The list of regions bound only by condition 1.
diffenrich2	The list of regions bound only by condition 2.
X	A n x p matrix of enrichment for each region and each condition, at the given FDR cutoff (1: enriched, 0: not-enriched).
diffprob1	A n-dimensional vector of posterior probabilities of differential binding for the two conditions under study; diffprob0=1-diffprob1.
diffX1	A n-dimensional index of regions bound only by condition 1 (0: not bound, 1: bound).
diffX2	A n-dimensional index of regions bound only by condition 2.
IPE	A p-dimensional vector of estimated IP efficiency values for each experiment.

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Joint modelling of ChIP-seq data via a Markov random field model, Biostatistics 2014, 15(2):296-310 DOI:10.1093/biostatistics/kxt047.

See Also

See also mrf, mrf.joint

FDR *Identification of enriched regions by controlling a given FDR level*

Description

Identify enriched regions by controlling false discovery rate at a specified level.

Usage

```
FDR(prob0, cr = 0.05)
```

Arguments

prob0	A numeric vector. The probability of the null hypothesis being true (i.e. a region not-enriched).
cr	A numeric variable. The level of FDR for identifying the enriched regions.

Value

X	The index of enriched regions (1: enriched, 0: not-enriched).

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Accounting for immunoprecipitation efficiencies in the statistical analysis of ChIP-seq data. BMC Bioinformatics 2013, 14:169 DOI:10.1186/1471-2105-14-169.

See Also

See also mix,mix.joint,enrich.mix

IPE *Estimating the ImmunoPrecipitation (IP) efficiency of a ChIP-seq experiment.*

Description

Calculate the IP efficiency of an experiment by using the mixture model parameters.

Usage

```
IPE(para, method = NULL)
```

Arguments

para	A numeric vector. The parameters estimated by the mixture model.
method	A charater variable. Can be "poisson" or "NB" and it refers to the densities of the mixture distribution.

Value

IPE estimated value.

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Accounting for immunoprecipitation efficiencies in the statistical analysis of ChIP-seq data. BMC Bioinformatics 2013, 14:169 DOI:10.1186/1471-2105-14-169.

See Also

See also `mix`,`mix.joint`,`enrich.mix`

---

mix *Fitting mixture of two densities, either Poisson or Negative Binomial, to ChIP-seq data.*

Description

`mix` uses an EM algorithm to fit ChIP-seq count data by a latent mixture model with two components. One component is the signal density and the other is the background density. `mix` can deal with more than one experiment at the same time. In this case, it fits individual models to each experiment. The output of this function can be used for further analysis by `mix.joint` or `enrich.mix`.

Usage

```
mix(data, method = NULL, initialpara=NULL, fixoffset=FALSE, fixk=3,krange=c(0:10),
     exp.label=NULL, stopdiff=1e-04, parallel=FALSE)
```

Arguments

data	A list, whose first argument is a n x 3 matrix with information on the bins. The three columns should contain "Chromosome", "Start" and "Stop" information. The second list contains the counts of ChIP-seq experiments. This is a n x p matrix, where n is the number of bins and p is the number of experiments. Count data for at least one experiment should be given.
method	A character variable. Can be "Poisson" or "NB" and it refers to the densities of the mixture distribution.

initialpara	A numeric matrix or vector. The initial parameters given for EM algorithm. In form of c("p", "lambda_S", "lambda_B") if method="Poisson" or c("p", "mu_S", "phi_S", "mu_B", "phi_B") if method="NB". Could be a matrix if initial values are the different for multiple experiments or a vector if initial values are the same. If not given, then a default value of (0.1, 10, 1) or (0.1, 10, 1, 1, 1) for method="Poisson" or "NB" respectively.
fixoffset	A logical variable. If TRUE, the offset of the signal distribution is fixed by the user and is the same for all experiments. If FALSE, the offset is estimated empirically for each experiment. Default value is FALSE.
fixk	A numeric variable. The value of the offset, when fixoffset = TRUE.
krange	A numeric vector. The range of the offset, when fixoffset = FALSE. Default range is from 0 to 10.
exp.label	A charater vector, giving a label for each experiments.
stopdiff	A numeric variable. A prescribed small quantity for determining the convergence of the EM algorithm. Default value is 1e-04.
parallel	A logical variable. If TRUE, then the individual experiments will be processed in parallel, using the clusterApplyLB function in package parallel. Default value is TRUE.

Value

data	The data provided as input.
parameters	The parameters estimated by the mixture model. The parameters are (p, lambda_S, lambda_B, k) when method="Poisson" or (p, mu_S, phi_S, mu_B, phi_B, k) when method="NB". p is the proportion of signal in the mixture model. For a Poisson mixture model, lambda_S and lambda_B represent the mean of the signal and mean of the background, respectively. For a NB mixture model, mu_S and phi_S are the mean and overdispersion of the signal density, respectively, whereas mu_B and phi_B are the mean and overdispersion of the background density, respectively.
method	The method used for the analysis

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Accounting for immunoprecipitation efficiencies in the statistical analysis of ChIP-seq data. BMC Bioinformatics 2013, 14:169 DOI:10.1186/1471-2105-14-169.

See Also

See also mix.joint, enrich.mix

Examples

```
tempdir()
data(p300cbp.1000bp)
exp.label=c("CBPT0", "CBPT301", "CBPT302", "p300T0",
"p300T301", "p300T302", "WangCBP", "Wangp300")
## Simple examples -- only two experiments and first 5000 observations
CBPT30=list()
CBPT30$region=p300cbp.1000bp$region[1:5000,]
CBPT30$count=p300cbp.1000bp$count[1:5000,2:3]
Poissonfit.simple<-mix(CBPT30, method="Poisson", exp.label=exp.label[c(2,3)])
```

---

mix.joint *Joint fitting of mixture of Poisson or NB densities to ChIP-seq experiments.*

Description

`mix.joint` uses an EM algorithm to jointly fit ChIP-seq data for two or more experiments. Technical replicates are accounted for in the model as well as individual ChIP efficiencies for each experiment. Prior biological knowledge, such as the expectation of a similar number of binding profiles for the same protein under two similar conditions, can also be included in the model to aid robustness in the detection of enriched and differentially bound regions. The output of `mix.joint` can be further analysed by `enrich.mix`.

Usage

```
mix.joint(data, method = NULL, para.sep = NULL, rep.vec = NULL,
    p.vec = NULL, exp.label = NULL, stopdiff = 1e-04)
```

Arguments

`data`	A list, whose first argument is a n x 3 matrix with information on the regions. The three columns should contain "Chromosome", "Start" and "Stop" information. The second list contains the counts of ChIP-seq experiments. This is a n x p matrix, where n is the number of regions and p is the number of experiments. Count data for at least one experiment should be given.
`method`	A character variable. Can be "Poisson" or "NB" and it refers to the densities of the mixture distribution.
`para.sep`	A p x q matrix, where p is the number of experiments and q is the number of parameters in the mixture model. This is used as initial parameters for the joint modelling function. We recommend using the parameters of `mix`, as these are optimized for each experiment. If there are no technical replicates, then the parameters of the mix function are automatically used for the mix.joint output.
`rep.vec`	A non-zero integer vector. The vector of replicate indices, of length equal to the number of experiments. Technical replicates share the same index, e.g c(1,2,2,3,4,4,5,6) for 8 experiments where the 2nd and 3rd are two technical replicates and similarly the 5th and 6th.

p.vec	A non-zero integer vector. The vector of p indices, with p=P(X_s=1) for any region s. This vector is of length equal to the number of experiments. Experiments with the same probability of enrichment share the same p index, such as technical replicates and/or proteins with a similar number of binding sites, e.g. c(1,1,1,2,3,3,3,4) if the first three experiments have the same p and similarly the 5th, 6th and 7th experiments. This allows to propertly account for the different IP efficiencies in the joint analysis. At least one of rep.vec or p.vec should be given. For those experiments which do not share the same index (p.vec or rep.vec) with any other experiment, a single mixture model will be fitted.
exp.label	A charater vector, giving the labels for each experiment.
stopdiff	A numeric variable. A prescribed small quantity for determining the convergence of the EM algorithm. Default value is 1e-04.

Value

data	The data provided as input.
parameters	The parameters estimated by the mixture model. The parameters are (p, lambda_S, lambda_B, k) when method="poisson" or (p, mu_S, phi_S, mu_B, phi_B, k) when method="NB". p is the proportion of signal in the mixture model. For a Poisson mixture model, lambda_S and lambda_B represent the mean of the signal and mean of the background, respectively. For a NB mixture model, mu_S and phi_S are the mean and overdispersion of the signal density, respectively, whereas mu_B and phi_B are the mean and overdispersion of the background density, respectively.
rep.vec	The rep.vec used for the analysis
p.vec	The p.vec used for the analysis
method	The method used for the analysis

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Accounting for immunoprecipitation efficiencies in the statistical analysis of ChIP-seq data. BMC Bioinformatics 2013, 14:169 DOI:10.1186/1471-2105-14-169.

See Also

See also mix, enrich.mix

Examples

```
tempdir()
data(p300cbp.1000bp)
exp.label=c("CBPT0", "CBPT301", "CBPT302", "p300T0",
    "p300T301", "p300T302", "WangCBP", "Wangp300")
## Simple examples -- only two experiments and first 5000 observations
```

```
CBPT30=list()
CBPT30$region=p300cbp.1000bp$region[1:5000,]
CBPT30$count=p300cbp.1000bp$count[1:5000,2:3]
Poisfit.simple<-mix(CBPT30, method="Poisson", exp.label=exp.label[c(2,3)])
## Joint analysis combining technical replicates
##   (CBPT301,CBPT302)
Poisfit.joint<-mix.joint(CBPT30, Poisfit.simple$parameters, method="Poisson",
    rep.vec=c(1,1), p.vec=c(1,1), exp.label=exp.label[c(2,3)])
```

`mrf` *Fitting a one-dimensional Markov random field mixture model to ChIP-seq data.*

Description

`mrf` uses an MCMC algorithm to fit a one-dimensional Markov random field model for the latent binding profile from ChIP-seq data. The emission distribution of the enriched state (signal) can be either Poisson or Negative Binomial (NB), while the emission distribution of the non-enriched state (background) can be either a Zero-inflated Poisson (ZIP) or a Zero-inflated Negative Binomial (ZINB).

Usage

```
mrf(data, method=NULL, exp.label = NULL, Niterations=10000, Nburnin=5000,
     Poisprior=c(5, 1, 0.5, 1), NBprior=c(5, 1, 1, 1, 0.5, 1, 1, 1),
    PoisNBprior=c(5,1,1,1, 0.5,1), var.NB=c(0.1, 0.1, 0.1, 0.1), parallel=TRUE)
```

Arguments

`data`	A list, whose first argument is a n x 3 matrix with information on the bins. The three columns should contain "Chromosome", "Start" and "Stop" information. The second argument contains the counts of a single ChIP-seq experiment. This is a n x 1 matrix, where n is the number of bins.
`method`	A character variable. Can be "Poisson", "PoisNB" or "NB" and it refers to the densities of the mixture distribution. "Poisson" means that a ZIP distribution is used for the background (with parameters pi and mean lambda_B) and a Poisson distribution for the signal (with parameter lambda_S); "PoisNB" means that a ZIP distribution is used for the bacground (with parameter pi and lambda_B) and a NB distribution for the signal (with mean mu_S and overdispersion phi_S); "NB" means that a ZINB distribution is used for the background (with parameters pi, mu_B and phi_B) and a NB distribution for the signal (with mean mu_S and overdispersion phi_S).
`exp.label`	A charater vector, giving a label for experiment.
`Niterations`	An integer value, giving the number of MCMC iteration steps. Default value is 10000.
`Nburnin`	An integer value, giving the number of burn-in steps. Default value is 5000.

`Poisprior` The gamma priors for the parameter lambda in the Poisson-Poisson mixture: the first two elements are the priors for signal and the second two are priors for background. Default values are (5,1, 0.5, 1).

`NBprior` The gamma priors for the mean mu and overdispersion parameter phi in the NB-NB mixture: the first two elements are the priors for mu_S for the signal; the third and fourth elements are priors for phi_S; the fifth and sixth elements are priors for mu_B for the background and the seventh and eighth are priors for phi_B. Default values are (5, 1, 1, 1, 0.5, 1, 1, 1).

`PoisNBprior` The gamma priors for lambda_B and mu_S, phi_S in Poisson-NB mixture, the first two are priors for mu_S, the third and the fourth are priors for phi_S, the fifth and the sixth are priors for lambda_B. Default values are (5, 1,1,1, 0.5, 1).

`var.NB` The variances used in the Metropolis-Hastling algorithm for estimating (mu_S, phi_S, mu_B, phi_B) for NB mixture or for estimating (mu_S, phi_S) for PoisNB mixture. Default values are (0.1, 0.1, 0.1, 0.1) or (0.1, 0.1) for NB and PoisNB respectively.

`parallel` A logical variable. If TRUE and the experiment has more than one chromosome, then the individual chromosomes will be processed in parallel, using the `clusterApplyLB` function in package `parallel`. Default value is TRUE.

Value

`data` The data provided as input.

`parameters` The estimates of parameters which are the mean of samples of parameters.

`parameters.sample`
The samples matrix drawing from the posterior distributions of the parameters. The samples are collected one from every ten steps right after burn-in step. The column names for the matrix are (q_1, q_0, lambda_S, pi, lambda_B) if method="Poisson" or (q_1, q_0, mu_S, phi_S, pi, mu_B, phi_B) if method ="NB" or (q_1, q_0, mu_S, phi_S, pi, lambda_B) if method="PoisNB", where q_1 and q_0 are the transition probabilities that the current bin is enriched given the previous bin is enriched or not enriched, respectively.

`PP` The posterior probabilities that bins are enriched.

`method` The method used for the analysis.

`acrate.NB` The acceptance rate of Metropolis-Hastling method.

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Joint modelling of ChIP-seq data via a Markov random field model, Biostatistics 2014, 15(2):296-310 DOI:10.1093/biostatistics/kxt047.

See Also

#See also `mrf.joint`,`enrich.mrf`

`mrf.joint` *Joint fitting of a one-dimensional Markov random field model to multiple ChIP-seq datasets.*

Description

`mrf.joint` uses an MCMC algorithm to fit one-dimensional Markov random field models to multiple ChIP-seq datasets. These datasets could contain technical and biological replicates. If a single experiment is given, then the function `mrf` is used. The emission distribution of the enriched state (signal) could be either Poisson or Negative Binomial (NB), while the emission distribution of the non-enriched state (background) could be either a Zero-inflated Poisson (ZIP) or a Zero-inflated Negative Binomial (ZINB).

Usage

```
mrf.joint(data, method = NULL, rep.vec = NULL, p.vec = NULL, exp.label = NULL,
      Niterations = 10000, Nburnin = 5000, Poisprior = NULL, NBprior = NULL,
      PoisNBprior = NULL, var.NB = NULL, var.q=NULL, parallel=FALSE)
```

Arguments

`data`	A list, whose first argument is a n x 3 matrix with information on the regions. The three columns should contain "Chromosome", "Start" and "Stop" information. The second list contains the counts of ChIP-seq experiments. This is a n x p matrix, where n is the number of regions and p is the number of experiments. Count data for at least one experiment should be given.
`method`	A character variable. Can be "Poisson", "PoisNB" or "NB" and it refers to the densities of the mixture distribution. "Poisson" means that a ZIP distribution is used for the background (with parameters pi and mean lambda_B) and a Poisson distribution for the signal (with parameter lambda_S); "PoisNB" means that a ZIP distribution is used for the bacground (with parameter pi and lambda_B) and a NB distribution for the signal (with mean mu_S and overdispersion phi_S) ; "NB" means that a ZINB distribution is used for the background (with parameters pi, mu_B and phi_B) and a NB distribution for the signal (with mean mu_S and overdispersion phi_S).
`rep.vec`	A non-zero integer vector. The vector of replicate indices, of length equal to the number of experiments. Technical replicates share the same index, e.g c(1,2,2,3,4,4,5,6) for 8 experiments where the 2nd and 3rd are two technical replicates and similarly the 5th and 6th.
`p.vec`	A non-zero integer vector. The vector of p indices, with p=P(X_s=1) for any region s. This vector is of length equal to the number of experiments. Experiments with the same probability of enrichment share the same p index, such as technical replicates and/or proteins with a similar number of binding sites, e.g. c(1,1,1,2,3,3,3,4) if the first three experiments have the same p and similarly the 5th, 6th and 7th experiments. This allows to propertly account for the different IP efficiencies in the joint analysis. At least one of rep.vec or p.vec should

be given. For those experiments which do not share the same index (p.vec or rep.vec) with any other experiment, function `mrf` will be used.

`exp.label` A charater vector, giving a label for each experiments.

`Niterations` An integer value, giving the number of MCMC iteration step. Default value is 10000.

`Nburnin` An integer value, giving the number of burn-in step. Default value is 5000.

`Poisprior` The gamma priors for the parameter lambda in the Poisson-Poisson mixture: the first two elements are the priors for signal and the second two are priors for background. If p experiments are given, then the prior should be a matrix 4 x p, where each column represents the priors for each experiment. Default values are (5,1, 0.5, 1) for each single experiment.

`NBprior` The gamma priors for the mean mu and overdispersion parameter phi in the NB-NB mixture: the first two elements are the priors for mu_S for the signal; the third and fourth elements are priors for phi_S; the fifth and sixth elements are priors for mu_B for the background and the seventh and eighth are priors for phi_B. If p experiments are given then the prior should be a matrix 8 x p, where each column represents the priors for each experiment. Default values are (5, 1, 1, 1, 0.5, 1, 1, 1) for each single experiment.

`PoisNBprior` The gamma priors for lambda_B and mu_S, phi_S in Poisson-NB mixture, the first two are priors for mu_S, the third and the fourth are priors for phi_S, the fifth and the sixth are priors for lambda_B. If more than one experiments are given then the prior should be a matrix 6 x p, each column represents the priors for each experiment. Default values are (5, 1,1,1, 0.5, 1) for each single experiment.

`var.NB` The variances used in Metropolis-Hastling algorithm for estimates of (mu_S, phi_S, mu_B, phi_B) for NB mixture or for estimates of (mu_S, phi_S) for PoisNB mixture. If p experiments are given then var.NB should be 4 x p or 2 x p matrix for NB and PoisNB respectively, each column represents the variances used for each experiment. Default values for each single experiment are (0.1, 0.1, 0.1, 0.1) or (0.1, 0.1) for NB and PoisNB respectively.

`var.q` the variances used in Metropolis-Hastling algorithm for estimates of q_0 and common ratio parameter when assume same p condition for multiple experiments. The number of components of var.q equals to number of experiment +1. Default values are 0.001 for each experiment and 0.005 for common ratio parameter. For example, var.q=(0.001, 0.001, 0.005) for two experiments.

`parallel` A logical variable. If TRUE and the experiment has more than one chromosome, then the individual chromosomes will be processed in parallel, using the `clusterApplyLB` function in package `parallel`. Default value is TRUE.

Value

`data` The data provided as input.

`parameters` The list of parameters for each experiment, where each list contains the samples matrix drawing from the posterior distributions of the parameters. The samples are collected one from every ten steps right after burn-in step. The column names for the matrix are (q_1, q_0, lambda_S, pi, lambda_B) if method="Poisson"

	or (q_1, q_0, mu_S, phi_S, pi, mu_B, phi_B) if method ="NB" or (q_1, q_0, mu_S, phi_S, pi, lambda_B) if method="PoisNB", where q_1 and q_0 are the transition probabilities that the current region is enriched given the previous region is enriched or not enriched, respectively.
PP	The list of posterior probabilities for each experiment, where each list contains a vector of posterior probabilities that regions are enriched.
rep.vec	The rep.vec used for the analysis.
p.vec	The p.vec used for the analysis.
method	The method used for the analysis.
acrate.NB	The acceptance rate of Metropolis-Hastling method.

Author(s)

Yanchun Bao and Veronica Vinciotti

References

Bao et al. Joint modelling of ChIP-seq data via a Markov random field model, Biostatistics 2014, 15(2):296-310 DOI:10.1093/biostatistics/kxt047.

See Also

#See also `mrf.joint`,`enrich.mrf`

`p300cbp.1000bp` *Example Data*

Description

`p300cbp.1000bp` contains ChIP-seq counts in 1000bp length bins on chromosome 21 for 8 experiments. The names of the 8 experiments are CBPT0, CBPT301, CBPT302, p300T0, p300T301, p300T302, WangCBP, Wangp300. The data contain two lists, first list is the region information which contains 3 columns: Chromosome, Start, Stop and the second list are the count.

Usage

```
data(p300cbp.1000bp)
```

Format

List of 2:
region:'data.frame', 33916 obs. of 3 variables, Chromosome, Start, Stop
count: 'numeric', 33916 obs. of 8 variables, CBPT0, CBPT301, CBPT302, p300T0, p300T301, p300T302, WangCBP, Wangp300

Source

The first 6 datasets, CBPT0, CBPT301, CBPT302, p300T0, p300T301, p300T302, are from the GEO database, accession number GSE21026.

The last 2 datasets, WangCBP, Wangp300, are from the GEO database, accession number GSE15735.

References

Ramos et al. (2010) Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. Nucleic Acids Research, 38(16):5396-5408.

Wang et al. (2009) Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. Cell, 138:1019-1031.

Examples

```
data(p300cbp.1000bp)
```

`p300cbp.200bp` *Example Data*

Description

`p300cbp.200bp` contains ChIP-seq counts in 200bp length bins on chromosome 21 for 8 experiments. The names of the 8 experiments are CBPT0, CBPT301, CBPT302, p300T0, p300T301, p300T302, WangCBP, Wangp300. The data are consistu

Usage

```
data(p300cbp.200bp)
```

Format

List of 2:
region:'data.frame', 234721 obs. of 3 variables, Chromosome, Start, Stop
count: 'numeric', 234721 obs. of 8 variables, CBPT0, CBPT301, CBPT302, p300T0, p300T301, p300T302, WangCBP, Wangp300

Source

The first 6 datasets are from the GEO database, accession number GSE21026.

The last 2 data sets are from the GEO database, accession number GSE15735.

References

Romas, et.al, 2010. Genome-wide assessment of differential roles for p300 and CBP in transcription regulation. Nucleic Acids Research, 38(16):5396-5408.

Wang, et.al, 2009. Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes. Cell, 138:1019-1031.

Examples

```
data(p300cbp.200bp)
```

Index