Package 'BDMMAcorrect'

April 15, 2020

Type Package

Title Meta-analysis for the metagenomic read counts data from different cohorts

Version 1.4.0

Author ZHENWEI DAI <daizwhao@gmail.com>

Maintainer ZHENWEI DAI <daizwhao@gmail.com>

Date 2018-10-27

Description Metagenomic sequencing techniques enable quantitative analyses of the microbiome. However, combining the microbial data from these experiments is challenging due to the variations between experiments. The existing methods for correcting batch effects do not consider the interactions between variables—microbial taxa in microbial studies—and the overdispersion of the microbiome data. Therefore, they are not applicable to microbiome data. We develop a new method, Bayesian Dirichlet-multinomial regression meta-analysis (BDMMA), to simultaneously model the batch effects and detect the microbial taxa associated with phenotypes. BDMMA automatically models the dependence among microbial taxa and is robust to the high dimensionality of the microbiome and their association sparsity.

License GPL (>= 2)

Depends R (>= 3.5), vegan, ellipse, ggplot2, ape, SummarizedExperiment

Encoding UTF-8

LazyData true

Imports Rcpp (>= 0.12.12), RcppArmadillo, RcppEigen, stats

LinkingTo Rcpp, RcppArmadillo, RcppEigen

biocViews ImmunoOncology, BatchEffect, Microbiome, Bayesian

RoxygenNote 6.0.1

Suggests knitr, rmarkdown, BiocGenerics

VignetteBuilder knitr

git_url https://git.bioconductor.org/packages/BDMMAcorrect

git_branch RELEASE_3_10

git_last_commit 8359f9b

git_last_commit_date 2019-10-29

Date/Publication 2020-04-14

R topics documented:

`BDMMA` *Bayesian Dirichlet–Multinomial approach for meta-analysis of metagenomic read counts*

Description

Bayesian Dirichlet–Multinomial approach for meta-analysis of metagenomic read counts

Usage

```
BDMMA(Microbiome_dat, abundance_threshold = 5e-05, burn_in = 5000,
  sample_period = 5000, bFDR = 0.1, PIPcut = 0.5)
```

Arguments

`Microbiome_dat`	A SummarizedExperiment object that includes the taxonomy read counts, phenotypes and batch labels.
`abundance_threshold`	The minimum abundance level for the taxa to be included (default value = 5e-05).
`burn_in`	The length of burn in period before sampling the parameters (default value = 5,000).
`sample_period`	The length of sampling period for estimating parameters' distribution (default value = 5,000)
`bFDR`	The false discovery rate level to control (default value = 0.1).
`PIPcut`	The threshold to cut the posterior inclusion probabilities (PIPs). By default, PIP is thresholding at 0.5.

Value

A list contains the selected taxa and summary of parameters included in the model.

`selected.taxa`	A list includes the selected taxa fesatures that are significantly associated with the main effect variable.
`parameter_summary`	A data.frame contains the mean and quantiles of the parameters included in the model. Each row includes a parameter's distribution summary and the parameter name is labeled in the first row. alpha_g: the baseline intercept of g-th taxon; betaj_g: the association strength between the g-th taxon and j-th input variables; deltai_g: the batch effect parameter of batch i, taxon g; L_g: the posterior selection probability of g-th taxon; p: the proportion of significantly associated taxa; eta: the standard deviation of the spike distribution (in the spike-and-slab prior).

PIP	A vector contains the PIPs of selected microbial taxa.
bFDR	The corresponding bFDR under the selected microbial taxa.

References

Dai, Zhenwei, et al. "Batch Effects Correction for Microbiome Data with Dirichlet-multinomial Regression." Bioinformatics 1 (2018): 8.

Examples

```
require(SummarizedExperiment)
data(Microbiome_dat)
## (not run)
## output <- BDMMA(Microbiome_dat, burn_in = 3000, sample_period = 3000)
```

---

fdr_cut *Threshold the posterior inclusion probability (PIP) through control Bayesian false discovery rate (bFDR).*

Description

Threshold the posterior inclusion probability (PIP) through control Bayesian false discovery rate (bFDR).

Usage

```
fdr_cut(PIP_vec, alpha = 0.1)
```

Arguments

PIP_vec	A vector contains the PIPs of parameters
alpha	The level of the bFDR to need to control (default = 0.1)

Value

The cutoff for PIPs to control the bFDR with the user defined value, alpha.

Examples

```
data(L_mean)
cutoff <- fdr_cut(L_mean, alpha = 0.1)
```

L_mean *Posterior Inclusion Probabilities (PIP)*

Description

A dataset containing the posterior inclusion probabilities of 40 variables

Usage

```
L_mean
```

Format

A numeric vector including 40 PIP values

Microbiome_dat *Taxonomy Reads and Associated Phenotypes*

Description

Simualated taxonomy read counts of 40 taxa and their associated phenotypes.

Usage

```
Microbiome_dat
```

Format

SummarizedExperiment

Details

The dataset contains the simualated taxonomy read counts from 80 samples, where the samples come from 4 different batches and include both case and control samples in each batch. For the detailed usuage, please see the package vignette.

trace_plot *Trace plot of BDMMA output*

Description

Trace plot of BDMMA output

Usage

```
trace_plot(trace, param, col = "black")
```

Arguments

trace	A data.frame named "trace" contained in the output of function BDMMA
param	A character vector including the parameters' name for trace_plot
col	A string defining the color of trace plot (default color is black)

Value

The function returns a list containing plot objects of parameters' trace plot.

Examples

```
require(SummarizedExperiment)
data(Microbiome_dat)
## (not run)
## output <- BDMMA(Microbiome_dat, burn_in = 3000, sample_period = 3000)
## figure <- trace_plot(output$trace, param = c("alpha_1", "beta1_10"))
## print(figure)
```

VBatch *Visualize batch effect with principal coordinate analysis*

Description

Visualize batch effect with principal coordinate analysis

Usage

```
VBatch(Microbiome_dat, main_variable = NULL, method = "bray")
```

Arguments

Microbiome_dat	A SummarizedExperiment object that includes the taxonomy read counts, phenotypes and batch labels.
main_variable	Optional. A vector containing the main effect variable. Only for categorical main effect variable.The function will generate a figure for each catagory.
method	A string indicating which method should be used to calculate the distance matrix for principal coordinate analysis.

Value

The function returns a list containing plot objects of principal coordinate analysis figures.

Examples

```
data(Microbiome_dat)
figure <- VBatch(Microbiome_dat, method = "bray")
print(figure)
```

Index