Package ‘BDMMACorrect’

May 16, 2019

Type Package

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

Version 1.2.0

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_9

git_last_commit 7ca38db

git_last_commit_date 2019-05-02

Date/Publication 2019-05-16
R topics documented:

- BDMMA ................................. 2
- fdr_cut .................................. 3
- L_mean .................................. 4
- Microbiome_dat ....................... 4
- trace_plot ............................... 5
- VBatch .................................. 5

Index 7

| 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 features 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; deltag_i_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).
\textbf{fdr_cut}

- \textbf{PIP}\hspace{1cm} A vector contains the PIPs of selected microbial taxa.
- \textbf{bFDR}\hspace{1cm} The corresponding bFDR under the selected microbial taxa.

\textbf{References}


\textbf{Examples}

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

\textbf{fdr_cut}\hspace{1cm} \textit{Threshold the posterior inclusion probability (PIP) through control Bayesian false discovery rate (bFDR).}

\textbf{Description}

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

\textbf{Usage}

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

\textbf{Arguments}

- \textbf{PIP_vec}\hspace{1cm} A vector contains the PIPs of parameters
- \textbf{alpha}\hspace{1cm} The level of the bFDR to need to control (default = 0.1)

\textbf{Value}

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

\textbf{Examples}

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

---

**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**

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

**Usage**

Microbiome_dat

**Format**

SummarizedExperiment

**Details**

The dataset contains the simulated 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 usage, please see the package vignette.
trace_plot

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

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, phe-
notypes 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 category.
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

«Topic datasets
 L_mean, 4
 Microbiome_dat, 4

BDOMMA, 2

fdr_cut, 3

L_mean, 4

Microbiome_dat, 4

trace_plot, 5

VBatch, 5