Package 'RCRnorm'

January 20, 2025

Title An Integrated Regression Model for Normalizing 'NanoString nCounter' Data

Version 0.0.2

Author Gaoxiang Jia [aut, cre],
Guanghua Xiao [aut],
Xinlei Wang [aut]

Maintainer Gaoxiang Jia <GJia@SMU.edu>

Description 'NanoString nCounter' is a medium-throughput platform that measures gene or microRNA expression levels. Here is a publication that introduces this platform: Malkov (2009) <doi:10.1186/1756-0500-2-80>. Here is the webpage of 'NanoString nCounter' where you can find detailed information about this platform <https://www.nanostring.com/scientific-content/technology-overview/ncounter-technology>. It has great clinical application, such as diagnosis and prognosis of cancer. Implements integrated system of random-coefficient hierarchical regression model to normalize data from 'NanoString nCounter' platform so that noise from various sources can be removed.

Depends R (>= 2.15.0), truncnorm

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

NeedsCompilation no

Repository CRAN

Date/Publication 2018-02-22 22:00:03 UTC

Contents

FFPE_dat *FFPE data on 83 regular genes and 28 patients.*

Description

Data from lung cancer patients.

Usage

```
data(FFPE_dat)
```

Format

An object of class "list".

References

publication to be added (PubMed)

Examples

```
data(FFPE_dat)
```

RCRnorm *An Integrated Regression Model for Normalizing 'NanoString nCounter' Data*

Description

'NanoString nCounter' is a medium-throughput platform that measures gene or microRNA expression levels. Here is a publication that introduces this platform: Malkov (2009) <doi:10.1186/1756-0500-2-80>. Here is the webpage of NanoString nCounter where you can find detailed information about this platform <https://www.nanostring.com/scientific-content/technology-overview/ncounter-technology>. It has great clinical application, such as diagnosis and prognosis of cancer. This function implements an integrated system of random-coefficient hierarchical regression model for normalizing 'NanoString nCounter' data. It removes noise from the data so that expression levels of genes can be compared across patients.

Usage

```
RCRnorm(dat, pos_conc = log10(c(128, 32, 8, 2, 0.5, 0.125)),
  fast_method = FALSE, iter = 8000, warmup = 5000, random_init = F,
  all_dat = T, seed = 1, mm = 3, m_ab = 9)
```

Arguments

dat	A list containing data for the 4 probe types: positive control, negative control, housekeeping gene and regular gene. The names for the 4 elements in the list should exactly be: pos_dat, neg_dat, hk_dat and reg_dat, respectively. For an example of the input data format, please refer to the FFPE_dat included in the dataset. The data for each probe type should be a dataframe with rows being genes and column being patients. The number of columns (patients) should be the same for data of all four probe types. The rows of positive control data should have the same order as the postive control RNA amount vector supplied to the function.
pos_conc	A vector of log10 RNA amount of the positive controls. The order of these controls should be the same as the rows of positive control data in dat. The defaut is: log10(c(128, 32, 8, 2, 0.5, 0.125)).
fast_method	Logical flag; set to FALSE by default; when set to TRUE, the algorithm will implement a very fast method to estimate the normalized gene expression levels. It will first estimate sample specific slope and intercept from positive controls and then get the RNA levels of regular genes with the intercepts and slopes. Then two way anova will be performed on the RNA levels. The residuals from two way anova where sample effects and gene effects are removed will be the normalized expression levels of regular genes.
iter	Total number of iterations for Monte Carlo simulation. Default is 8000.
warmup	Number of burnin cycles for Monte Carlo simulation. Default is 5000.
random_init	Whether to estimate the starting point from data
all_dat	Whether should all data be used to update a_i and b_i.
seed	Seed for the MCMC sampling for reproducibility. Default is 1.
mm	Number of standard deviations for the prior uniform range.
m_ab	Number of variance for the prior distribution of mu_a and mu_b.

Details

'NanoString nCounter' platform includes several internal controls (Positive control; Negative control; Housekeeping genes) to remove noise and normalize data to enable inter-patient gene expression comparasion: 1. removing lane-by-lane experimental variation with positive controls; 2. removing background noise introduced by none specific binding with negative controls; 3. removing sample loading amount variation or difference in RNA degradation level with housekeeping genes. Our IBMnorm model integrates information from these 3 types of internal controls and get the normalized expression levels of genes we are interested in. Detailed models are in the publication.

Value

The function returns a list of elements including: summary statistics of key parameters in the model and a list of MCMC samples. The number of MCMC samples equals iter-warmup. If fast_method flag is set to TRUE, only normalized expression level matrix of regular genes will be returned with each column being a sample and each row being a gene.

Examples

```
data(FFPE_dat)
result = RCRnorm(FFPE_dat, iter = 20, warmup = 0)
```

Index