Package 'multiDE'

November 4, 2016

Type Package

Title Differential expression (DE) analysis for RNA-seq data with multiple treatment conditions

Version 1.0

Date 2016-04-17

Author Guangliang Kang, Hong Zhang

Maintainer Hong Zhang <zhangh@fudan.edu.cn>

Depends R (>= 2.10), edgeR (>= 3.12.1)

Suggests DESeq2 (>= 1.10.1), S4Vectors (>= 0.8.11)

Description multiDE is an R package for identifying differentially expressed genes between matched or unmatched samples under two or more treatment conditions.

License GPL (>= 2)

RoxygenNote 5.0.1

URL http://github.com/zhanghfd/multiDE

BugReports http://github.com/zhanghfd/multiDE/issues

R topics documented:

`multiDE-package` *Differential expression (DE) analysis for RNA-seq data with multiple treatment conditions*

Description

This package provides a statistical method for identifying DE genes using RNA-seq read count data from multiple conditions. Each condition has at least two samples, which can be either unmatched or matchced. The efficiency of this method is much improved over the existing methods through the use of a dimension reduced ANOVA model.

Details

Package:	multiDE
Type:	Package
Version:	1.0
Date:	2016-04-17
License:	GPL (>= 2)

Author(s)

Guangliang Kang and Hong Zhang

Maintainer: Hong Zhang <zhangh@fudan.edu.cn>

References

Kang G., Du L., Zhang H. (2016). multiDE: A dimension reduced model based statistical method for differential expression analysis using RNA-sequencing data with multiple treatment conditions. BMC Bioinformatics 67:248.

`cell` *cellular phenotype data*

Description

a study of Homo sapiens hormone embryonic stem cells was downloaded from the NCBI GEO database (accession ID: GSE36552). To find the casual relationship between gene expression network and cellular phenotype, Yan et al. derived embryonic stem cells from donated human preimplantation embryos, prepared cDNA and sequenced them by Illumina HiSeq2000.

Usage

```
data(cell)
```

Format

A data frame with 6,526 observations on the following 10 variables.

`nameOfGene` Genes name

`2cell_e1` Read count for sample 1 of 2-cell stage

`2cell_e2` Read count for sample 2 of 2-cell stage

`2cell_e3` Read count for sample 3 of 2-cell stage

`4cell_e1` Read count for sample 1 of 4-cell stage

`4cell_e2` Read count for sample 2 of 4-cell stage

`4cell_e3` Read count for sample 3 of 4-cell stage

`8cell_e1` Read count for sample 1 of 8-cell stage

`8cell_e2` Read count for sample 2 of 8-cell stage

`8cell_e3` Read count for sample 3 of 8-cell stage

References

Yan L., Yang M., Guo H., Yang L., Wu J., Li R., Liu P., Lian Y., Zheng X., Yan J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. Nature Structural & Molecular Biology 20(9), 1131-1139.

`dispersion` *Esimate both common and tagwise dispersion parameters for a dataset with a prescribed experimental design.*

Description

This function provides dispersion parameter estimates. The imputs are normalized RNA-seq read count data, an experimental design vector and a logical variable indicating whether the samples are matched or not. The output is a list containing the estimated common dispersion "commonDispersion" and the estimated tagwise dispersion vector "tagwiseDispersion".

Usage

```
dispersion(M,condition,matched,method="edgeR")
```

Arguments

`M`	A list generated by the function "normalization", which contains the normalized read counts matrix and the estimated size factor.
`condition`	A vector indicating the condition statuses of all samples.
`matched`	A logical variable indicating whether the samples from various conditions are matched (TRUE) or not (FALSE).
`method`	The method used to estimate the dispersion parameters. This should be either "edgeR" or "DESeq2". The default normalization method is "edgeR".

Value

count	Normalized read counts matrix, rows for genes and columns for samples.
sizeFactor	Estimated size factor.
condition	A vector indicating the condition statuses of all samples.
commonDispersion	Estimated common dispersion parameter.
tagwiseDispersion	Estimated tagwise dispersion parameter.
matched	A logical variable indicating whether the samples from various conditions are matched (TRUE) or not (FALSE).

Author(s)

Guangliang Kang and Hong Zhang

References

Kang G, Du L, Zhang H (2016). multiDE: A dimension reduced model based statistical method for differential expression analysis using RNA-sequencing data with multiple treatment conditions.

Examples

```
data(psoriatic);

count = as.matrix(psoriatic[,-1]);

M = normalization(count,method="median");

condition = c(rep(1:8,each=3));

M = dispersion(M,condition=condition,matched=TRUE,method="edgeR");
```

`multiDE` *Differential expression analysis for RNA-seq data with multiple treatment conditions*

Description

This function provides DE analysis results. The imput is a list generated by the function "dispersion", and the outputs include p-values, fold changes, and the variances of fold change for each gene.

Usage

```
multiDE(M,n.top=nrow(M$count))
```

Arguments

M	A list generated by the function "dispersion", which contains estimated dispersion parameters, condition statuses, and a logical variable indicating whether the samples are matched or not.
n.top	The number of genes used for estimating parameter vector u, which should be smaller or equal to the number of all genes. The default value is the number of all genes.

Value

log2FoldChange	log2-fold changes.
p.value	P-values for DE analysis.
u	Estimation of parameters u.
sd.log2FoldChange	Standard errors of estimated log2-fold changes.

Author(s)

Guangliang Kang and Hong Zhang

References

Kang G, Du L, Zhang H (2016). multiDE: A dimension reduced model based statistical method for differential expression analysis using RNA-sequencing data with multiple treatment conditions.

Examples

```
data(cell);

count = as.matrix(cell[,-1]);

M = normalization(count,method="median");

M = dispersion(M,condition=rep(1:3,each=3),matched=FALSE,method="edgeR");

result = multiDE(M,n.top=1e3);
```

---

normalization *Normalize the raw RNA-seq read count data.*

Description

This function normalizes raw count data. The imputs include a raw count data matrix and the normalization method. There are four options of normalizations methods, namely meandian normalization (median), total normalization (total), quantile normalization (quantile), and trimmed mean of M-values (TMM). The outputs is a list containing the estimated size factor and normalized count matrix.

Usage

```
normalization(count,method="median")
```

Arguments

count	Raw RNA-seq read count data matrix, rows for genes and columns for samples.
method	The normalization method to be used. This should be one of "median", "total", "quantile", or "TMM". The default normalization method is "median".

Value

count	Normalized read counts matrix, rows for genes and columns for samples.
sizeFactor	Estimated size factor.

Author(s)

Guangliang Kang and Hong Zhang

References

Kang G, Du L, Zhang H (2016). multiDE: A dimension reduced model based statistical method for differential expression analysis using RNA-sequencing data with multiple treatment conditions.

Examples

```
data(psoriatic);

count = as.matrix(psoriatic[,-1]);

M = normalization(count,method="median");
```

---

psoriatic *psoriatic study data*

Description

In this study, the major interest was to detect the influence of aryl hydrocarbon receptor (AhR) on RNA expression profiles of psoriatic lesion cells. Each of eight patients were treated with culture treatment of DMSO (vehicle control), AhR agonist FICZ, and AhR antagonist CH-2233191. RNA-seq data were obtained using Illumina Genome Analyzer II platform for each of three treated lesion tissue samples. Therefore, this was a matched sample design. The RNA-seq read counts data were derived from the GEO database (accession ID: GSE47944). There were altogether 13,416 genes with maximal read counts greater than 50 in each treatment condition.

Usage

```
data(psoriatic)
```

Format

A data frame with 13,416 observations on the following 25 variables.

`Gene` Genes name

`Sample_PDM_K1_2` Read count for patient 1 in vehicle control group

`Sample_PDM_K2_2` Read count for patient 2 in vehicle control group

`Sample_PDM_K3_2` Read count for patient 3 in vehicle control group

`Sample_K4.2` Read count for patient 4 in vehicle control group

`Sample_K5.2` Read count for patient 5 in vehicle control group

`Sample_K6_2` Read count for patient 6 in vehicle control group

`Sample_K8_2` Read count for patient 7 in vehicle control group

`Sample_K9_2` Read count for patient 8 in vehicle control group

`Sample_PDM_K1_3` Read count for patient 1 in AhR agonist group

`Sample_PDM_K2_3` Read count for patient 2 in AhR agonist group

`Sample_PDM_K3_3` Read count for patient 3 in AhR agonist group

`Sample_K4.3` Read count for patient 4 in AhR agonist group

`Sample_K5.3` Read count for patient 5 in AhR agonist group

`Sample_K6_3` Read count for patient 6 in AhR agonist group

`Sample_K8_3` Read count for patient 7 in AhR agonist group

`Sample_K9_3` Read count for patient 8 in AhR agonist group

`Sample_PDM_K1_4` Read count for patient 1 in AhR antagonist group

`Sample_PDM_K2_4` Read count for patient 2 in AhR antagonist group

`Sample_PDM_K3_4` Read count for patient 3 in AhR antagonist group

`Sample_K4.4` Read count for patient 4 in AhR antagonist group

`Sample_K5.4` Read count for patient 5 in AhR antagonist group

`Sample_K6_4` Read count for patient 6 in AhR antagonist group

`Sample_K8_4` Read count for patient 7 in AhR antagonist group

`Sample_K9_4` Read count for patient 8 in AhR antagonist group

References

Di Meglio P, Duarte J.H., Ahlfors H., Owens N.D., Li Y., Villanova F., Tosi I., Hirota K., Nestle F.O., Mrowietz U., et al. (2014). Activation of the aryl hydrocarbon receptor dampens the severity of in ammatory skin conditions. Immunity 40(6), 989-1001.

Index