Package 'FUCHIKOMA'

December 9, 2015

Type Package

Title Detection of Differentially Expressed Genes in one of multiple clusters using BAHSIC and Diffusion Map

Version 0.0.1

Date 2015-12-3

Author Koki Tsuyuzaki, Haruka Ozaki, Mika Yoshimura, Itoshi Nikaido

Maintainer Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

Depends R (>= 2.5.0), destiny, doParallel, foreach

biocViews Bioinformatics, DifferentialExpression, Clustering, MultipleComparisons, RNAseq, HighThroughputSequencing, Sequencing, Software

Description FUCHIKOMA detects differentially expressed genes (DEGs) in one of multiple clusters. To detect DEGs, FUCHIKOMA has two calculation mode; ``supervised-mode" and ``unsupervised-mode". In supervised-mode, FUCHIKOMA detects DEGs by using a label vector, in which the cluster of each sample or cell is written. In unsupervised-mode, FUCHIKOMA detects DEGs without the label vector. In this mode, user run the diffusion map, and specify which diffusion components contribute to the difference of such cluster.

License Artistic-2.0

NeedsCompilation no

R topics documented:

FUCHIKOMA-package *Detection of Differentially Expressed Genes in one of multiple clusters using BAHSIC and Diffusion Map*

Description

FUCHIKOMA detects differentially expressed genes (DEGs) in one of multiple clusters. To detect DEGs, FUCHIKOMA has two calculation mode; "supervised-mode" and "unsupervised-mode". In supervised-mode, FUCHIKOMA detects DEGs by using a label vector, in which the cluster of each sample or cell is written. In unsupervised-mode, FUCHIKOMA detects DEGs without the label vector. In this mode, user run the diffusion map, and specify which diffusion components contribute to the difference of such cluster.

Details

The DESCRIPTION file: This package was not yet installed at build time.

Index: This package was not yet installed at build time.
The main function is FUCHIKOMA, which returns an object containing the calculation results.

Author(s)

Koki Tsuyuzaki, Haruka Ozaki, Mika Yoshimura, Itoshi Nikaido

Maintainer: Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

References

Koki Tsuyuzaki et al. (2015) Fuchikoma: Detection of Differentially Expressed Genes in one of multiple clusters using BAHSIC and Diffusion Map. R package version 1.0.0

Laleh Haghverdi et al. (2015) Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics, 31(18), 2989-2998

Le Song et al. (2007) Gene selection via the BAHSIC family of algorithms, Bioinformatics, 23(13), i490-i498

Y-h Taguchi et al. (2015) Principal component analysis-based unsupervised feature extraction applied to in silico drug discovery for posttraumatic stress disorder-mediated heart disease, BMC Bioinformatics, 16(139)

Diego Adhemar Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Naama Elefant, Franziska Paul, Irina Zaretsky, Alexander Mildner, Nadav Cohen, Steffen Jung, Amos Tanay, Ido Amit (2014) Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science*, **343 (6172)**: 776-779

Arthur Gretton et al. (2007) A Kernel Statistical Test of Independence, NIPS 21

See Also

DiffusionMap

`CatKernel` *Categorical Kernel function*

Description

CatKernel calculate gram matrix by using label information (e.g., cell label or sample label such as c(1,1,1,2,2,2,3,3)).

Usage

```
CatKernel(label, type = c("two", "one_vs_rest", "each", "simple"))
```

Arguments

`label`	Label vector descrbing which cells are sample cell type and which cells are different (e.g., c(1,1,1,2,2,2,3,3)).
`type`	Categorical kernel functions. simple could be specified even the number of class is two or more. When the number of class is two, only two parameter is accessible. When the number of class is three or more, one_vs_rest and each is accessible.

Value

A data frame is returned.

Author(s)

Koki Tsuyuzaki, Haruka Ozaki, Mika Yoshimura, Itoshi Nikaido

Maintainer: Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

References

Le Song et al. (2007) Gene selection via the BAHSIC family of algorithms, Bioinformatics, 23(13), i490-i498

Y-h Taguchi et al. (2015) Principal component analysis-based unsupervised feature extraction applied to in silico drug discovery for posttraumatic stress disorder-mediated heart disease, BMC Bioinformatics, 16(139)

Examples

```
data(MARS)
L <- CatKernel(label.MARS, type="one_vs_rest")
dim(L)
```

`custom.DiffusionMap` *Customized Diffusion Map of destiny for extracting Gram Matrix*

Description

Inner function of FUCHIKOMA, which returns the gram matrix of diffusion map.

Usage

```
custom.DiffusionMap(data, sigma = NULL, k = find.dm.k(nrow(data) - 1L), n.eigs = min(20L, nrow(dat
```

Arguments

`data`	Expression data to be analyzed. Provide `vars` to select specific columns other than the default: all double value columns
`sigma`	Diffusion scale parameter of the Gaussian kernel. Either a number or a Sigmas object. (Optional. default: will be calculated using find.sigmas) A larger sigma might be necessary if the eigenvalues can not be found because of a singularity in the matrix
`k`	Number of nearest neighbors to consider (default: a guess betweeen 100 and $n - 1$) `NULL` or `NA` are also interpreted as `n - 1L`.
`n.eigs`	Number of eigenvectors/values to return (default: 20)
`density.norm`	logical. If TRUE, use density normalisation
`...`	All parameter after this are optional and have to be specified by name
`distance`	Distance measurement method. Euclidean distance (default) or cosine distance ($1 - corr(c_1, c_2)$).
`censor.val`	Value regarded as uncertain. Either a single value or one for every dimension (Optional, default: CENSOR.VAL)
`censor.range`	Uncertainity range for censoring (Optional, default: none). A length-2-vector of certainty range start and end. TODO: also allow $2 \times G$ matrix
`missing.range`	Whole data range for missing value model. Has to be specified if NAs are in the data
`vars`	Variables (columns) of the data to use. Specifying NULL will select all columns (default: All floating point value columns)
`verbose`	Show a progressbar and other progress information (default: do it if censoring is enabled)
`.debug.env`	If supplied, the rotated transition probaility matrix `M` and the scaling rotation `D.rot` is stored in it prior to eigen decomposition.

Value

A DiffusionMap object:

Author(s)

Koki Tsuyuzaki, Haruka Ozaki, Mika Yoshimura, Itoshi Nikaido

Maintainer: Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

References

Laleh Haghverdi et al. (2015) Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics, 31(18), 2989-2998

See Also

DiffusionMap

FUCHIKOMA *Detection of Differentially Expressed Genes in one of multiple clusters using BAHSIC and Diffusion Map*

Description

This package was not yet installed at build time.

Usage

```
FUCHIKOMA(data, mode = c("Supervised", "Unsupervised"), Comp = FALSE, label = FALSE, cat.type = FAl
```

Arguments

data	A data matrix, in which the row means genes and column means cells or sample.
mode	When Supervised is specified, FUCHIKOMA uses label paramter. When Unsupervised is specified, FUCHIKOMA uses Comp parameter for specifying which diffusion components should be used.
Comp	When mode is specified as Unsupervised, Comp must be specified such as c(1,2).
label	When mode is specified as Supervised, label must be specified such as c(1,1,1,2,2,2,3,3)
cat.type	Type of categorical kernel of CatKernel
n.eigs	Number of eigenvectors/values to return (default: 20)
algorithm	brute means single gene rejection strategies and song means fixed percent of genes rejection strategies in each iteration step of FUCHIKOMA.
per.rej	When algorithm is specified as song, per.rej must be specified such as 20 (default:10)
threshold	In each iteration step, if the difference of HSIC in the step and max value of previous HSICs is lower than threshold, iteration will be halted (default: 0.01).

Value

DEGs.HSICs	Differentially expressed genes and HSIC values
DEGs.Pvals	Pvalues of HSICs
All.HSICs	HSICs value of all genes
Rej.order	The order of rejecting of genes in each iteration step

Author(s)

Koki Tsuyuzaki, Haruka Ozaki, Mika Yoshimura, Itoshi Nikaido

Maintainer: Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

References

Laleh Haghverdi et al. (2015) Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics, 31(18), 2989-2998

Le Song et al. (2007) Gene selection via the BAHSIC family of algorithms, Bioinformatics, 23(13), i490-i498

Y-h Taguchi et al. (2015) Principal component analysis-based unsupervised feature extraction applied to in silico drug discovery for posttraumatic stress disorder-mediated heart disease, BMC Bioinformatics, 16(139)

See Also

DiffusionMap

Examples

```
data(MARS)
res <- FUCHIKOMA(data=MARS, mode="Supervised", label=label.MARS, type="one_vs_rest", n.eigs=10, algorithm="s
```

HSIC *Hilbert-Schmidt Independence Criteria (HSIC)*

Description

Inner function of FUCHIKOMA. HSIC calculated aginst two gram-matrix means the independence of two kernel spaces. The higher HSIC value is, the more two space are dependent. When p.value set TRUE, p-value of HSIC is also calculated by moment matching to a gamma distribution.

Usage

```
HSIC(K, L, p.value = FALSE)
```

Arguments

K	First gram matrix
L	Second gram matrix
p.value	When p.value is specified as TRUE, p-value of HSIC is calculated (default:FALSE)

Value

HSIC	The value of HSIC
Pval	P-value of HSIC. The value is accessible only when p.value is specified as TRUE

Author(s)

Koki Tsuyuzaki, Haruka Ozaki, Mika Yoshimura, Itoshi Nikaido

Maintainer: Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

References

Le Song et al. (2007) Gene selection via the BAHSIC family of algorithms, Bioinformatics, 23(13), i490-i498

Y-h Taguchi et al. (2015) Principal component analysis-based unsupervised feature extraction applied to in silico drug discovery for posttraumatic stress disorder-mediated heart disease, BMC Bioinformatics, 16(139)

Arthur Gretton et al. (2007) A Kernel Statistical Test of Independence, NIPS 21

Examples

```
K <- matrix(runif(100), nrow=10)
L <- matrix(runif(100), nrow=10)
HSIC(K, L, p.value)
```

MARS *Count data of MARS-Seq containing four different cell types.*

Description

A data frame with 10801 rows (genes) with following 228 columns (cells).

The data is downloaded form GEO (GSE54006) and genes having no expression are filted.

The data has four different cell types; B cells (B1-B48), Dendritic cells (DC1-DC89), Monocytes (Mono1-Mono46), and Natural Killer cells (NK1-NK45).

Usage

```
data(MARS)
```

Author(s)

Koki Tsuyuzaki, Haruka Ozaki, Mika Yoshimura, Itoshi Nikaido

Maintainer: Koki Tsuyuzaki <k.t.the-answer@hotmail.co.jp>

Source

https://www.sciencemag.org/content/343/6172/776?related-urls=yes&legid=sci;343/6172/776

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE54006

References

Diego Adhemar Jaitin, Ephraim Kenigsberg, Hadas Keren-Shaul, Naama Elefant, Franziska Paul, Irina Zaretsky, Alexander Mildner, Nadav Cohen, Steffen Jung, Amos Tanay, Ido Amit (2014) Massively Parallel Single-Cell RNA-Seq for Marker-Free Decomposition of Tissues into Cell Types. *Science*, **343 (6172)**: 776-779

Examples

```
data(MARS)
dim(MARS)
pairs(result.pca.MARS$rotation[,1:10], col=label.MARS, main="MARS-Seq (PCA)")
pairs(result.destiny.MARS@eigenvectors[,1:10], col=label.MARS, main="MARS-Seq (Diffusion Map)")
```

Index