Package 'oposSOM'

July 2, 2014

Type Package

Title Package for analysis and visualization of metagenes

Version 0.2-3

Date 2013-08-13

Author Henry Wirth

Maintainer Henry Wirth <wirth@izbi.uni-leipzig.de>

Description This package translates microarray expression data into metagenes and provides various visualizations and second level analyses.

License GPL (>= 2)

LazyLoad yes

URL http://som.izbi.uni-leipzig.de

Depends R (>= 2.11.0)

Imports som, fastICA, pixmap, igraph, biomaRt

NeedsCompilation no

Repository CRAN

Date/Publication 2013-08-13 12:18:57

R topics documented:

oposSOM-package *Package for analysis and visualization of metagenes*

Description

This package translates microarray expression data into metagenes and provides various visualizations and second level analyses. For a given data set, a standardized analysis pipeline is processed and the output is stored in an external folder. Additionally, a HTML summary file is created collating and linking all analysis results.

Details

Package:	oposSOM
Type:	Package
Version:	0.2-3
Date:	2013-08-13
License:	GPL (>= 2)
LazyLoad:	yes

Author(s)

Henry Wirth

Maintainer: Henry Wirth <wirth@izbi.uni-leipzig.de>

References

Wirth, Loeffler, v.Bergen, Binder: Expression cartography of human tissues using self organizing maps. (BMC Bioinformatics 2011)

Wirth, v.Bergen, Binder: MALDI-typing of infectious algae of the genus Prototheca using SOM portraits. (Journal of microbiological methods 2012)

Examples

```
library( oposSOM )


#### Toy example ####

## create data
toy.data = matrix( runif(10000), 1000, 10 )

## run pipeline
run.pipeline( toy.data, dataset.name = "Example", dim.som1 = 20, dim.som2 = 20 )
```

```
#### Real data example - takes some time to process ####

## load data
#data( tissues )

## run pipeline
#run.pipeline( tissue.data, dataset.name="Tissue Example" )
```

`run.pipeline` *Run the oposSOM pipeline.*

Description

This function realizes the complete pipeline functionality: single gene expression values are cultered to metagenes using a self-organizing map. Based on these metagenes, visualizations (e.g. expression profiles) and downstreaming analysis approaches (e.g. hierarchical clustering, ICA or geneset overrepresentation analysis) are performed.

Usage

```
run.pipeline( indata, dataset.name = "Unnamed", dim.som1 = 20, dim.som2 = 20 )
```

Arguments

`indata`	numerical data matrix containing genes as rows and samples as columns.
`dataset.name`	name of the folder which will be created to contain the results.
`dim.som1`	size of the primary SOM / resolution of the expression profiles: number of tiles equals dim.som1^2.
`dim.som2`	resolution of the second level SOM: number of tiles equals dim.som2^2.

Details

If `indata` is supplied with ENSEMBL-IDs as rownames, geneset overrepresentation analysis is carried out using predefined GO gene sets automatically downloaded from Ensembl database (http://www.ensembl.org). The results are contained in summary sheets of the integrated maps, e.g. the overexpression overview map.

Value

The pipeline will create a folder in the working directory according to `dataset.name` containing all results. Graphical output is provided in respective files sole to ensure functionality in command line mode.

Additionally, a HTML summary file is created which contains brief descriptions and links to all results:

- Expression profiles for all samples in different color modes(standard, loglog and WAD).

- Supporting maps illustrating properties of the SOM as metagene population, variance and concordance.

- Second level SOM providing sample-centered view on relations and similarities between the samples.

- Metagene based downstreaming analysis: Hierarchical clustering, pairwise correlation matrix and spanning tree, and independent component analysis.

- Report sheets for the spot summary maps containing various information of identified clusters of metagenes.

Author(s)

Henry Wirth

References

Wirth, Loeffler, v.Bergen, Binder: Expression cartography of human tissues using self organizing maps. (BMC Bioinformatics 2011)

Wirth, v.Bergen, Binder: MALDI-typing of infectious algae of the genus Prototheca using SOM portraits. (Journal of microbiological methods 2012)

Examples

```
library( oposSOM )


#### Toy example ####

## create data
toy.data = matrix( runif(10000), 1000, 10 )

## run pipeline
run.pipeline( toy.data, dataset.name = "Example", dim.som1 = 20, dim.som2 = 20 )


#### Real data example - takes some time to process ####

## load data
#data( tissues )

## run pipeline
#run.pipeline( tissue.data, dataset.name="Tissue Example" )
```

`tissue.data` *Example data set.*

Description

A data set comprising of 12 selected human tissues.

Usage

```
data(tissues)
```

Format

The data set is stored in RData (binary) format.

Details

The data set was downloaded from Gene Expression Omnibus repository (http://www.ncbi.nlm.nih.gov/geo, GEO accession no. GSE7307). About 38,000 genes in more than 650 samples were measured using the Affymetrix HGU133-Plus2 microarray. A subset of 12 selected tissues from different categories is used as example data set for the oposSOM-package. ENSEMBL-IDs are used as row-identifier to enable geneset overrepresentation analysis.

Source

http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7307

Index