Package ‘SCORPION’

September 24, 2025

Type Package

Title Single Cell Oriented Reconstruction of PANDA Individual Optimized Networks

Version 1.2.0

Description Constructs gene regulatory networks from single-cell gene expression data using the PANDA (Passing Attributes between Networks for Data Assimilation) algorithm.

License GPL-3

Encoding UTF-8

LazyData true

Imports cli, methods, irlba, igraph, RANN, Matrix, pbapply

Suggests RhpcBLASctl, gpuR

RoxygenNote 7.3.2

NeedsCompilation no

Author Daniel Osorio [aut, cre] (ORCID: <`https://orcid.org/0000-0003-4424-8422`>), Marieke L. Kuijjer [aut] (ORCID: <`https://orcid.org/0000-0001-6280-3130`>)

Maintainer Daniel Osorio <`daniecos@uio.no`>

Repository CRAN

Date/Publication 2025-09-24 21:50:02 UTC

Contents

`scorpion` *Constructs PANDA gene regulatory networks from single-cell gene expression data*

Description

Constructs gene regulatory networks from single-cell gene expression data using the PANDA (Passing Attributes between Networks for Data Assimilation) algorithm.

Usage

```
scorpion(
  tfMotifs = NULL,
  gexMatrix,
  ppiNet = NULL,
  computingEngine = "cpu",
  nCores = 1,
  gammaValue = 10,
  nPC = 25,
  assocMethod = "pearson",
  alphaValue = 0.1,
  hammingValue = 0.001,
  nIter = Inf,
  outNet = c("regulatory", "coregulatory", "cooperative"),
  zScaling = TRUE,
  showProgress = TRUE,
  randomizationMethod = "None",
  scaleByPresent = FALSE,
  filterExpr = FALSE
)
```

Arguments

`tfMotifs`	A motif dataset, a data.frame or a matrix containing 3 columns. Each row describes an motif associated with a transcription factor (column 1) a gene (column 2) and a score (column 3) for the motif.
`gexMatrix`	An expression dataset, with genes in the rows and barcodes (cells) in the columns.
`ppiNet`	A Protein-Protein-Interaction dataset, a data.frame or matrix containing 3 columns. Each row describes a protein-protein interaction between transcription factor 1(column 1), transcription factor 2 (column 2) and a score (column 3) for the interaction.
`computingEngine`	Either 'cpu' or 'gpu'
`nCores`	Number of processors to be used if BLAS or MPI is active.
`gammaValue`	Graining level of data (proportion of number of single cells in the initial dataset to the number of super-cells in the final dataset)

`nPC`	Number of principal components to use for construction of single-cell kNN network.
`assocMethod`	Association method. Must be one of ’pearson’, ’spearman’ or ’pcNet’.
`alphaValue`	Value to be used for update variable.
`hammingValue`	Value at which to terminate the process based on Hamming distance.
`nIter`	Sets the maximum number of iterations PANDA can run before exiting.
`outNet`	A vector containing which networks to return. Options include "regulatory", "coregulatory", "cooperative".
`zScaling`	Boolean to indicate use of Z-Scores in output. False will use [0,1] scale.
`showProgress`	Boolean to indicate printing of output for algorithm progress.
`randomizationMethod`	Method by which to randomize gene expression matrix. Default "None". Must be one of "None", "within.gene", "by.genes". "within.gene" randomization scrambles each row of the gene expression matrix, "by.gene" scrambles gene labels.
`scaleByPresent`	Boolean to indicate scaling of correlations by percentage of positive samples.
`filterExpr`	Boolean to indicate wheter or not to remove genes with 0 expression across all cells from the GEX input.

Value

A list of matrices describing networks achieved by convergence with PANDA algorithm.

Author(s)

Daniel Osorio <daniecos@uio.no>

Examples

```
# Loading example data
data(scorpionTest)

# The structure of the data
str(scorpionTest)

# List of 3
# $ gex:Formal class 'dgCMatrix' [package "Matrix”] with 6 slots
# .. ..@ i       : int [1:4456] 1 5 8 11 22 30 33 34 36 38 ...
# .. ..@ p       : int [1:81] 0 47 99 149 205 258 306 342 387 423 ...
# .. ..@ Dim     : int [1:2] 230 80
# .. ..@ Dimnames:List of 2
# .. .. ..$ : chr [1:230] "MS4A1” "CD79B” "CD79A” "HLA-DRA” ...
# .. .. ..$ : chr [1:80] "ATGCCAGAACGACT” "CATGGCCTGTGCAT” "GAACCTGATGAACC” "TGACTGGATTCTCA” ...
# .. ..@ x       : num [1:4456] 1 1 3 1 1 4 1 5 1 1 ...
# .. ..@ factors : list()
# $ tf :'data.frame': 4485 obs. of  3 variables:
#   ..$ tf    : chr [1:4485] "ADNP” "ADNP” "ADNP” "AEBP2” ...
# ..$ target: chr [1:4485] "PRF1” "TMEM40” "TNFRSF1B” "CFP” ...
# ..$ mor   : num [1:4485] 1 1 1 1 1 1 1 1 1 1 ...
```

```
# $ ppi:'data.frame': 12754 obs. of  3 variables:
#   ..$ X.node1       : chr [1:12754] "ADNP" "ADNP" "ADNP" "AEBP2" ...
# ..$ node2         : chr [1:12754] "ZBTB14" "NFIA" "CDC5L" "YY1" ...
# ..$ combined_score: num [1:12754] 0.769 0.64 0.581 0.597 0.54 0.753 0.659 0.548 0.59 0.654 ...

# Running SCORPION with large alphaValue for testing purposes.
scorpionOutput <- scorpion(tfMotifs = scorpionTest$tf,
                           gexMatrix = scorpionTest$gex,
                           ppiNet = scorpionTest$ppi,
                           alphaValue = 0.8)

# -- SCORPION -------------------------------------------------------------------------------------
# + Initializing and validating
# + Verified sufficient samples
# i Normalizing networks
# i Learning Network
# i Using tanimoto similarity
# + Successfully ran SCORPION on 214 Genes and 783 TFs

# Structure of the output.
str(scorpionOutput)

# List of 6
# $ regNet  :Formal class 'dgeMatrix' [package "Matrix"] with 4 slots
# .. ..@ x       : num [1:167562] -0.413 1.517 -1.311 0.364 -1.041 ...
# .. ..@ Dim     : int [1:2] 783 214
# .. ..@ Dimnames:List of 2
# .. .. ..$ : chr [1:783] "ADNP" "AEBP2" "AIRE" "ALX1" ...
# .. .. ..$ : chr [1:214] "ACAP1" "ACRBP" "ACSM3" "ADAR" ...
# .. ..@ factors : list()
# $ coregNet:Formal class 'dgeMatrix' [package "Matrix"] with 4 slots
# .. ..@ x       : num [1:45796] 7.07e+06 -4.06 1.76e+01 -1.16e+01 -1.62e+01 ...
# .. ..@ Dim     : int [1:2] 214 214
# .. ..@ Dimnames:List of 2
# .. .. ..$ : chr [1:214] "ACAP1" "ACRBP" "ACSM3" "ADAR" ...
# .. .. ..$ : chr [1:214] "ACAP1" "ACRBP" "ACSM3" "ADAR" ...
# .. ..@ factors : list()
# $ coopNet :Formal class 'dgeMatrix' [package "Matrix"] with 4 slots
# .. ..@ x       : num [1:613089] 5.65e+06 -5.16 -3.79 -3.63 2.94 ...
# .. ..@ Dim     : int [1:2] 783 783
# .. ..@ Dimnames:List of 2
# .. .. ..$ : chr [1:783] "ADNP" "AEBP2" "AIRE" "ALX1" ...
# .. .. ..$ : chr [1:783] "ADNP" "AEBP2" "AIRE" "ALX1" ...
# .. ..@ factors : list()
# $ numGenes: int 214
# $ numTFs  : int 783
# $ numEdges: int 167562
```

`scorpionTest` *Example single-cell gene expression, motif, and ppi data*

Description

This data is a list containing three objects. The motif `data.frame` describes a set of pairwise connections where a specific known sequence motif of a transcription factor was found upstream of the corresponding gene. The expression `dgCMatrix` is a set of 230 gene expression levels measured across 80 cells. Finally, the ppi `data.frame` describes a set of known pairwise protein-protein interactions.

Usage

```
data(scorpionTest)
```

Format

A list containing three datasets.

`gex` A subsetted version of 10X Genomics' 3k PBMC dataset provided by the `Seurat` package.

`tf` Subset of the transcription-factor and target gene list provided by the `dorothea` package for Homo sapiens.

`ppi` The known protein-protein interactions and the combined score downloaded from the STRING database

Examples

```
# Loading example data
data(scorpionTest)

# The structure of the data
str(scorpionTest)

# List of 3
# $ gex:Formal class 'dgCMatrix' [package "Matrix"] with 6 slots
# .. ..@ i       : int [1:4456] 1 5 8 11 22 30 33 34 36 38 ...
# .. ..@ p       : int [1:81] 0 47 99 149 205 258 306 342 387 423 ...
# .. ..@ Dim     : int [1:2] 230 80
# .. ..@ Dimnames:List of 2
# .. .. ..$ : chr [1:230] "MS4A1" "CD79B" "CD79A" "HLA-DRA" ...
# .. .. ..$ : chr [1:80] "ATGCCAGAACGACT" "CATGGCCTGTGCAT" "GAACCTGATGAACC" "TGACTGGATTCTCA" ...
# .. ..@ x       : num [1:4456] 1 1 3 1 1 4 1 5 1 1 ...
# .. ..@ factors : list()
# $ tf :'data.frame': 4485 obs. of  3 variables:
#   ..$ tf    : chr [1:4485] "ADNP" "ADNP" "ADNP" "AEBP2" ...
# ..$ target: chr [1:4485] "PRF1" "TMEM40" "TNFRSF1B" "CFP" ...
# ..$ mor   : num [1:4485] 1 1 1 1 1 1 1 1 1 1 ...
# $ ppi:'data.frame': 12754 obs. of  3 variables:
#   ..$ X.node1       : chr [1:12754] "ADNP" "ADNP" "ADNP" "AEBP2" ...
# ..$ node2         : chr [1:12754] "ZBTB14" "NFIA" "CDC5L" "YY1" ...
# ..$ combined_score: num [1:12754] 0.769 0.64 0.581 0.597 0.54 0.753 0.659 0.548 0.59 0.654 ...
```

Index