

Package ‘scINSIGHT’

January 24, 2022

Type Package

Title Interpretation of Heterogeneous Single-Cell Gene Expression Data

Version 0.1.2

Date 2022-01-19

Maintainer Kun Qian <Kun_Qian@foxmail.com>

Description We develop a novel matrix factorization tool named 'scINSIGHT' to jointly analyze multiple single-cell gene expression samples from biologically heterogeneous sources, such as different disease phases, treatment groups, or developmental stages. Given multiple gene expression samples from different biological conditions, 'scINSIGHT' simultaneously identifies common and condition-specific gene pathways and quantify their expression levels in each sample in a lower-dimensional space. With the factorized results, the inferred expression levels and memberships of common gene pathways can be used to cluster cells and detect cell identities, and the condition-specific gene pathways can help compare functional differences in transcriptomes from distinct conditions.

License GPL-3

Imports Rcpp, RANN, igraph, parallel, stats, stringr

LinkingTo Rcpp, RcppArmadillo

Depends methods

RoxygenNote 7.1.1

NeedsCompilation yes

Author Kun Qian [aut, ctb, cre] (<<https://orcid.org/0000-0002-2354-2238>>),
Wei Vivian Li [aut, ctb] (<<https://orcid.org/0000-0002-2087-2709>>)

Repository CRAN

Date/Publication 2022-01-24 14:12:43 UTC

R topics documented:

create_scINSIGHT	2
run_scINSIGHT	2
scINSIGHT-class	4

Index	5
--------------	----------

create_scINSIGHT *Create an scINSIGHT object.*

Description

This function initializes an scINSIGHT object with normalized data passed in.

Usage

```
create_scINSIGHT(norm.data, condition)
```

Arguments

norm.data	List of normalized expression matrices (genes by cells). Gene names should be the same in all matrices.
condition	Vector specifying sample conditions.

Value

scINSIGHT object with norm.data slot set.

Examples

```
# Demonstration using matrices with randomly generated numbers
S1 <- matrix(runif(50000,0,2), 500,100)
S2 <- matrix(runif(60000,0,2), 500,120)
S3 <- matrix(runif(80000,0,2), 500,160)
S4 <- matrix(runif(75000,0,2), 500,150)
data = list(S1, S2, S3, S4)
sample = c("sample1", "sample2", "sample3", "sample4")
condition = c("control", "activation", "control", "activation")
names(data) = sample
names(condition) = sample
scINSIGHTx <- create_scINSIGHT(data, condition)
```

run_scINSIGHT *Perform scINSIGHT on normalized datasets*

Description

Perform INterpreting single cell gene expresSIon bioloGically Heterogeneous daTa (scINSIGHT) to return factorized W_{ℓ_1} , W_{ℓ_2} , H and V matrices.

This factorization produces a W_{ℓ_1} matrix (cells by K_j), a W_{ℓ_2} matrix (cells by K), a shared V matrix (K by genes) for each sample, and a H (K_j by genes) matrix for each condition. W_{ℓ_2} are the expression matrices of K common gene pathways for all samples, V is the membership matrix of K common gene pathways, and it's shared by all samples. W_{ℓ_1} are the expression matrices of K_j condition-specific gene pathways for all samples, and H are the membership matrices of K_j condition-specific gene pathways for all conditions.

Usage

```
run_scINSIGHT(
  object,
  K = seq(5, 15, 2),
  K_j = 2,
  LDA = c(0.001, 0.01, 0.1, 1, 10),
  thre.niter = 500,
  thre.delta = 0.01,
  num.cores = 1,
  B = 5,
  out.dir = NULL,
  method = "increase"
)
```

Arguments

object	scINSIGHT object.
K	Number of common gene pathways. (default c(5, 7, 9, 11, 13, 15))
K_j	Number of dataset-specific gene pathways. (default 2)
LDA	Regularization parameters. (default c(0.001, 0.01, 0.1, 1, 10))
thre.niter	Maximum number of block coordinate descent iterations to perform. (default 500)
thre.delta	Stop iteration when the reduction of objective function is less than the threshold. (default 0.01)
num.cores	Number of cores used for optimizing factorizations in parallel (default 1).
B	Number of repeats with random seed from 1 to B. (default 5)
out.dir	Output directory of scINSIGHT results. (default NULL)
method	Method of updating the factorization (default "increase"). If provide multiple K , user can choose method between "increase" and "decrease". For "increase", the algorithm will first perform factorization with the least $K = K_1$. Then initialize $K_2 - K_1$ facotrs, where K_2 is the K sightly larger than K_1 , and perform facotrizatoin with these new facotrs. Continue this process until the largest K . For "decrease", the algorithm will first perform factorization with the largest $K = K_1$. Then choose K_2 facotrs, where K_2 is the K sightly less than K_1 , and perform facotrizatoin with these new facotrs. Continue this process until the least K .

Value

scINSIGHT object with W_1 , W_2 , H , V and parameters slots set.

`scINSIGHT-class`*The scINSIGHT Class*

Description

The scINSIGHT object is created from two or more single cell datasets. To construct a scINSIGHT object, the user needs to provide at least two normalized expression (or another single-cell modality) matrices and the condition vector.

Details

The key slots used in the scINSIGHT object are described below.

Slots

`norm.data` List of normalized expression matrices (genes by cells). Each matrix should have the same number and name of genes.

`condition` Vector specifying each sample's condition name.

`W_1` List of W_{ℓ_1} estimated by scINSIGHT, names correspond to sample names.

`W_2` List of W_{ℓ_2} estimated by scINSIGHT, names correspond to sample names.

`H` List of H estimated by scINSIGHT, names correspond to condition names.

`V` Matrix V estimated by scINSIGHT.

`norm.W_2` List of W_{ℓ_2} after normalization. Recommended for downstream analysis.

`parameters` List of selected parameters, including K and λ .

Index

`create_scINSIGHT`, [2](#)

`run_scINSIGHT`, [2](#)

`scINSIGHT` (`scINSIGHT-class`), [4](#)

`scINSIGHT-class`, [4](#)