

Package ‘SCINA’

July 18, 2019

Type Package

Title A Semi-Supervised Category Identification and Assignment Tool

Version 1.2.0

Date 2019-7-16

Author Ze Zhang

Maintainer Ze Zhang <Ze.Zhang@utsouthwestern.edu>

Depends R (>= 2.15.0), MASS, gplots

Description An automatic cell type detection and assignment algorithm for single cell RNA-Seq and Cytof/FACS data. 'SCINA' is capable of assigning cell type identities to a pool of cells profiled by scRNA-Seq or Cytof/FACS data with prior knowledge of markers, such as genes and protein symbols that are highly or lowly expressed in each category. See Zhang Z, et al (2019) <doi:10.3390/genes10070531> for more details.

URL <http://lce.biohpc.swmed.edu/scina/>

<https://github.com/jcao89757/SCINA>

License GPL-2

NeedsCompilation no

RoxygenNote 6.1.0

Repository CRAN

Date/Publication 2019-07-18 06:38:25 UTC

R topics documented:

plotheat.SCINA	2
preprocess.signatures	2
SCINA	3

`plotheat.SCINA` *A function to plot SCINA results in a heatmap.*

Description

A function to plot SCINA results in a heatmap.

Usage

```
plotheat.SCINA(exp, results, signatures)
```

Arguments

<code>exp</code>	See more details in SCINA
<code>results</code>	An output object returned from SCINA.
<code>signatures</code>	See more details in SCINA

Value

Plot a heatmap showing signature genes' expression level and SCINA predicted cell types.

`preprocess.signatures`
A function to convert signatures uploaded via .csv files to lists used by SCINA.

Description

A function to convert signatures uploaded via .csv files to lists used by SCINA.

Usage

```
preprocess.signatures(file_path)
```

Arguments

<code>file_path</code>	The path of the .csv file. The first row of the file should be cell type names. Each column is occupied by the signature genes/protein markers for the cell type in the first row. Please find more details in SCINA.
------------------------	---

Value

A list of signature gene lists as an input for SCINA.

SCINA

*A semi-supervised cell type identification and assignment tool.***Description**

An automatic cell type detection and assignment algorithm for single cell RNA-Seq (scRNA-seq) and Cytof/FACS data. See Zhang Z, et al (2019) <doi:10.3390/genes10070531> for more details.

Usage

```
SCINA(exp, signatures, max_iter = 100, convergence_n = 10, convergence_rate = 0.99,
      sensitivity_cutoff = 1, rm_overlap = 1, allow_unknown = 1, log_file = "SCINA.log")
```

Arguments

<code>exp</code>	A normalized matrix representing the gene expression levels. The log-transformation is suggested to avoid heavy-tailed datasets. Columns correspond to cells, rows correspond to genes or protein symbols.
<code>signatures</code>	A list contains multiple signature vectors. Each signature vector contains genes or protein symbols, representing the prior knowledge for one cell type.
<code>max_iter</code>	An integer > 0. Default is 100. Max iterations allowed for the EM algorithm.
<code>convergence_rate</code>	A float between 0 and 1. Default is 0.99. Percentage of cells for which the type assignment remains stable for the last n rounds.
<code>sensitivity_cutoff</code>	A float between 0 and 1. Default is 1. The cutoff to remove signatures whose cells types are deemed as non-existent at all in the data by the SCINA algorithm.
<code>rm_overlap</code>	A binary value, default 1 (TRUE), denotes that shared symbols between signature lists will be removed. If 0 (FALSE) then allows different cell types to share the same identifiers.
<code>allow_unknown</code>	A binary value, default 1 (TRUE). If 0 (FALSE) then no cell will be assigned to the 'unknown' category.
<code>convergence_n</code>	An integer > 0. Default is 10. Stop the SCINA algorithm if during the last n rounds of iterations, cell type assignment keeps steady above the convergence_rate.
<code>log_file</code>	A string names the record of the running status of the SCINA algorithm, default 'SCINA.log'.

Details

More detailed information can be found from our web server: <http://lce.biohpc.swmed.edu/scina/>.

For any symbols in signature list, if the cell type is identified with symbol X's low detection level, please specify the symbol as 'low_X'. The name for the list is the cell type.

Details for 'low_X' (take scRNA-Seqs as an example):

(a) There are 4 cell types, the first one highly express one gene A, and the other three lowly express the same gene. Then it is better to specify A as the high marker for cell type 1, but it is not a good idea to specify A as the low expression marker for cell type 2,3,4.

(b) There are 4 cell types, the first one lowly express one gene A, and the other three highly express the same gene. Then it is better to specify A as the low marker for cell type 1, but it is not a good idea to specify A as the high expression marker for cell type 2,3,4.

(c) There are 4 cell types, the first one lowly express one gene A, the second and third one moderately express gene A, and the last one highly express gene A. Then it is better to specify A as the low marker for cell type 1, and as the high expression marker for cell type 4.

(d) The same specification can be applied to protein markers in CyTOF analysis.

Small sensitivity_cutoff leads to more signatures to be removed, and 1 denotes that no signature is removed.

Value

cell_labels return a vector contains cell type mapping results for each cell.

probabilities return a probability matrix indicating the predicted probability for each cell belongs to each cell type respectively.

Examples

```
load(system.file('extdata','example_expmat.RData', package = "SCINA"))
load(system.file('extdata','example_signatures.RData', package = "SCINA"))
exp = exp_test$exp_data
results = SCINA(exp, signatures, max_iter = 120, convergence_n = 12,
  convergence_rate = 0.999, sensitivity_cutoff = 0.9)
table(exp_test$true_label, results$cell_labels)
```