'bimir' R package manual

Maintainer: Sora Yoon yoonsora@unist.ac.kr

Last updated: 2018-11-05

1. Introduction

'bimir' is an R package developed to generate biclusters of cell condition-specific microRNA targets from large mRNA transcriptome fold change (FC) matrix (a.k.a., FC table; it contains log2 FC values of 20,639 human genes under 5,158 experimental conditions). Besides the primary purpose of this package, it can be applied to create any constant biclusters allowing small number of noises using the progressive bicluster extension (PBE) algorithm. This package is downloadable from the Github repository (https://github.com/ysora/bimir) and will not be uploaded to CRAN or Bioconductor due to the large size of FC table (~140 MB)

2. Installation

1. Download the bimir source package ('bimir_1.0.tar.gz').
2. Install and load the package by typing following lines.
   - install.packages('devtools')
   - library(devtools)
   - install_github('unistbig/bimir')
   - library(bimir)

3. Quick start with an example

Let's generate biclusters of 'hsa-miR-1-3p' where its target genes are up-regulated under common experimental conditions. To do this, type following lines.

- FC = load_FCtable() # To load FC table
- MP = getMIRprofile(miRNA = 'hsa-miR-1-3p', FCtable = FC, FCcutoff = log2(1.3)) # To generate MIR profile.
- PBE_MERGE(MIR_profile = MP, mir = 'hsa-miR-1-3p', biclust.path = './', FCcutoff = log2(1.3))
- Then check the directory assigned to 'biclust.path' and see the biclusters (biclust_up_*.txt; fig. 1) and corresponding experimental condition (Experimental_condition_up_*.txt; fig. 2) and gene list (Targetlist_up_*.txt; fig. 3).
- **Down-bicluster:** To generate biclusters of targets commonly down-regulated under multiple conditions, just modify the 'FCcutoff' parameter to be negative value (e.g., -log2(1.3)).
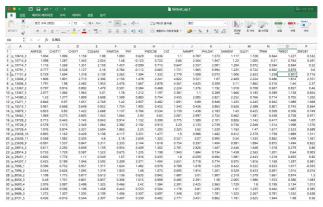- **The available miRNAs** are listed in the 'getmiRNAlist()' function.

Fig 1. Example bicluster showing log2-FC values of miR-1 targets under multiple conditions
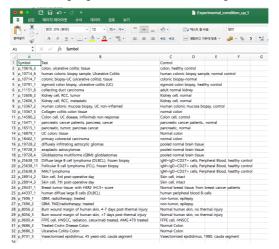

Fig 2. Experimental conditions (test and control conditions) for the example bicluster in figure 1

Fig 3. Gene list for the example bicluster in figure 1

4. Detailed function explanation

1. bicluster_extension_process

Extends biclusters allowing small noises

Usage: bicluster_extension_process(MIR_profile, seed_row, seed_col, zero_ratio)

Arguments

MIR_profile	A binarized microRNA profile
seed_row	List of a seed row index from MIR_profile
seed_col	List of a seed column index from MIR_profile
zero_ratio	Zero ratio to be allowed.

Value

Extended bicluster matrix

2. bicluster_trimming_process

Removes too noisy row or columns from bicluster.

Usage: bicluster_trimming_process(table, zero_ratio)

Arguments

table   A matrix that will be reduced

zero_ratio   Zero ratio allowed.

Value

Reduced bicluster matrix

3. **getGenelist**

Loads a vector of 20,639 human genes

Usage: getGenelist()

Value

A vector of 20,639 human genes

4. **getMIRprofile**

Returns MIR profile for input miRNA.

Usage: getMIRprofile(miRNA, FCtable = NULL, FCcutoff = log2(1.3))

Arguments

miRNA   Human mature miRNA name (e.g., hsa-miR-1-3p). Available miRNAs can be referred by typing 'getmiRNAlist()'

FCtable   FC matrix of 20,639 genes for 5,158 conditions. If NULL, it will be loaded from the local.

FCcutoff   Fold change cutoff in log2 scale. Default = log2(1.3). To extract MIR profile for target down-regulation in test condition, take negative value.

Value

A matrix of fold change value of sequence-based miRNA targets under selected cell conditions.

5. **get_miR_target**

Returns sequence-based miRNA targets predicted from three or more algorithms.

Usage: get_miR_target(miRNA)

Arguments

miRNA   Human mature miRNA name (e.g., hsa-miR-1-3p). Available miRNAs can be referred by typing 'getmiRNAlist()'

Value

A vector of sequence-based miRNA targets

6. get_zero_rate

Returns the ratio of zero (noise) of a binarized matrix

Usage: get_zero_rate(target)

Arguments

Target   a matrix

Value

The ratio of zero in a binarized matrix

7. getmiRNAlist

Loads human mature miRNA list

Usage: getmiRNAlist()

Value

A vector of 2,632 human mature miRNAs

8. load_FCtable

Loads a large table of fold change values for 20,639 human genes in 5,158 cell conditions.

Usage: load_FCtable()

Value

A fold change matrix of 5,158 rows and 20,639 columns.

9. make_biclust

Generates biclusters from MIR profile

Usage: make_biclust(MIR_PROFILE, FCcutoff, REPETITION, STEP_NUMBER, ZERORATE)

Arguments

MIR_PROFILE	A matrix of MIR profile.
FCcutoff	Binarization fold change cutoff in log2-scale for MIR_PROFILE. Target up-regulation biclusters will be created for positive FC cutoff, and down-regulation biclusters will be generated with negative FC cutoff.
REPETITION	The number of running ensemble bicluster function.
STEP_NUMBER	The number of extension process
ZERORATE	Final zero rate allowed.

Value

List of seed and biclusters

10. merge_bicluster

Merges similar biclusters

Usage: merge_bicluster(MIR_profile, tree_cutoff = 0.5, biclusts, seedlist,

FCcutoff = log2(1.3), mir = "testmiRNA", ZERORATE)

Arguments

MIR_profile	binarized microRNA profile
tree_cutoff	Similarity cutoff
Biclusts	list of extended biclusters
Seedlist	list of seeds
FCcutoff	Fold change cutoff (in log2 sclale) for binarization of MIR profile
Mir	Name of miRNA (e.g., hsa-miR-1-3p)
ZERORATE	Final zero ratio allowed

Value

Extended bicluster list

11. PBE

Progressive bicluster extension

Usage: PBE(MIR_profile, biclust_row, biclust_col, step_number, finalZR)

Arguments

MIR_profile	Binarized microRNA profile
biclust_row	List of index of row in seed biclust
biclust_col	List of index of columns in seed biclust
step_number	The number of extension process
finalZR	final zero rate allowed.

Value

List of row and column symbols of extended bicluster

12. PBE_MERGE

Progressive extension and merging of biclusters

Usage: PBE_MERGE(MIR_profile, mir, biclust.path = "./", FCcutoff = log2(1.3),

REPETITION = 10, STEP_NUMBER = 10, ZERORATE = 0.1, tree_cutoff = 0.5)

Arguments

MIR_profile	A matrix of microRNA profile where rows are cell conditions and columns are sequence-based miRNA targets.
mir	Name of microRNA
biclust.path	Directory where result biclusters and corresponding experimental condition and gene lists will be saved
FCcutoff	Log2-fold change binarization cutoff for seed. Default = log2(1.3)
REPETITION	The number of repetition that the ensemble function runs to extract seed biclusters. Default = 10
STEP_NUMBER	The number of extension process. Default = 10
ZERORATE	Final zero rate allowed. Default = 0.1
tree_cutoff	Similarity cutoff. Default = 0.5

Value

Extended and merged biclusters in the directory assigned to 'biclust.path'.