

'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

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

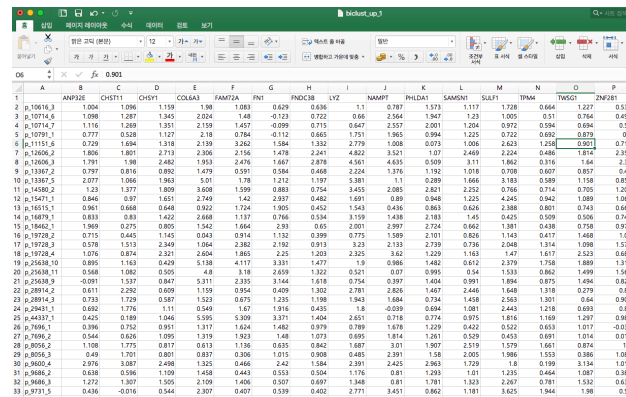


Fig 1. Example bicluster showing log2-FC values of miR-1 targets under multiple conditions

A	B	C	D	E	F	G
1	Symbol	Control				
2	p_10816.3	colon, ulcerative colitis; tissue	colon, healthy control			
3	p_10714.6	human colonic biopsy sample, Ulcerative Colitis	human colonic biopsy sample, normal control			
4	p_10714.7	colonic biopsy-UC, ulcerative colitis; tissue	colonic biopsy-normal			
5	p_10791.1	sigmoid colon biopsy, ulcerative colitis (UC)	sigmoid colon biopsy, healthy control			
6	p_11151.6	collecting duct carcinoma	adult normal kidney			
7	p_12606.2	Kidney cell, RCC, tumor	Kidney cell, normal			
8	p_12606.3	Kidney cell, RCC, metastatic	Kidney cell, normal			
9	p_13367.2	Human colonic mucosa biopsy, UC non-inflamed	Human colonic mucosa biopsy, control			
10	p_13367.5	Collagen colitis colon tissue	normal colon			
11	p_14580.2	Colon cell, UC disease, infliximab non-response	Colon cell, control			
12	p_15471.1	pancreatic cancer patients, pancreac; cancer	pancreatic cancer patients, normal			
13	p_16515.1	pancreatic, tumor; pancreas cancer	pancreatic, normal			
14	p_16879.1	UC colon tissue	Normal colon			
15	p_18462.1	primary colorectal carcinoma	normal colon			
16	p_19728.2	diffusely infiltrating astrocytic gliomas	pooled normal brain tissue			
17	p_19728.3	anaplastic astrocytomas	pooled normal brain tissue			
18	p_19728.4	Glioblastoma multiforme (GBM) glioblastomas	pooled normal brain tissue			
19	p_25638.10	Diffuse large B cell lymphoma (DLBCL), frozen biopsy	IgM+IgD+CD27+ cells, Peripheral Blood, healthy control			
20	p_25638.11	Follicular B cell lymphoma (FCL), frozen biopsy	IgM+IgD+CD27+ cells, Peripheral Blood, healthy control			
21	p_25638.9	MALT lymphoma	IgM+IgD+CD27+ cells, Peripheral Blood, healthy control			
22	p_28914.2	Skin cell, 3rd post-operative day	Skin cell, intact			
23	p_28914.3	Skin cell, 7th post-operative day	Skin cell, intact			
24	p_29431.1	Breast tumor tissue, with HER2, IHC3+ score	Normal breast tissue, from breast cancer patients			
25	p_44337.1	human diffuse large B cells (DLBCL)	human peripheral blood B-cells			
26	p_7696.1	GBM, radiotherapy, treated	non-tumor, epilepsy			
27	p_7696.2	GBM, TMZ/radiotherapy, treated	non-tumor, epilepsy			
28	p_8056.2	Burn wound margin of human skin, 4-7 days post-thermal injury	Normal human skin, no thermal injury			
29	p_8056.3	Burn wound margin of human skin, >7 days post-thermal injury	Normal human skin, no thermal injury			
30	p_9600.4	FFPE cell, HNSCC, radiation, cetuximab treated, AMG 479 treated	FFPE cell, HNSCC			
31	p_9686.2	Treated Crohn Disease Colon	Normal Colon			
32	p_9686.3	Ulcerative Colitis Colon	Normal Colon			
33	p_9731.5	Vasectomized epididymus, 45 years-old, cauda segment	Vasectomized epididymus, 1980, cauda segment			

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

Targetlist_up_1.txt

ANP32E
CHST11
CHSY1
COL6A3
FAM72A
FN1
FNDC3B
LYZ
NAMPT
PHLDA1
SAMSN1
SULF1
TPM4
TWSG1
ZNF281

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 scale) 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'.