Package 'FEM'

December 13, 2014

Type Package

Title Identification of FunctionalEpigenetic Modules

Version 1.0.0

Date 2014-02-17

Author Andrew E. Teschendorff and Yinming Jiao

Maintainer Andrew E. Teschendorff <andrew@picb.ac.cn>, Yinming Jiao <20907099@mail.zju.edu.cn>

Description FEM can dentify interactome hotspots of differential promoter methylation and differential ex-pression, where an inverse association between promoter methylation and gene expression is assumed.

License GPL (>=2)

Depends R (>= 2.10), Matrix, igraph, marray, corrplot, impute, limma, org.Hs.eg.db, graph, BiocGenerics

biocViews SystemsBiology,DNAMethylation,NetworkEnrichment,GeneRegulation,DifferentialMethylation,DifferentialExpression,N

R topics documented:

FEM-package *FEM*

Description

Identifies interactome hotspots ofdifferential promoter methylation and differentialexpression, where an inverse association between methylation and gene expression is assumed

Details

Package:	FEM
Type:	Package
Version:	1.0
Date:	2014-01-22
License:	GPL >2

Author(s)

Yinming Jiao and Andrew Teschendorff Maintainer: Yinming Jiao <20907099@zju.edu.cn>

References

1Jiao Y, Widschwendter M, Teschendorff AE. A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Bioinformatics 2014, doi: 10.1093/bioinformatics/btu316 (2014-05-02) 2 Jones A, Teschendorff AE, Li Q, Hayward JD, Kannan A, et al. (2013) Role of dna methylation and epigenetic silencing of hand2 in endometrial cancer development. PLoS Med 10:e1001551. 3 Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. Phys Rev E 74:016110. doi:10.1103/PhysRevE.74.016110. URL http://link.aps.org/doi/10.1103/PhysRevE.74.016110. 4 West J, Beck S, Wang X, Teschendorff AE (2013) An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. Sci Rep 3:1630.

DoEpiMod *DoEpiMod*

Description

Indentify differiential methylation hotspots in the network. Edge weights in the interactome network reflect the combined differiential methylation statistics(absolute values) of the genes making up the edge.

Usage

```
DoEpiMod(statM.m, adj.m, nseeds = 100, gamma = 0.5, nMC = 1000, sizeR.v = c(1, 100), minsizeOUT = 10, wri
```

Arguments

	Arguments:
	a matrix of statistics and p-values of differential methylation (one row for each gene promoter) with rownames annotated with entrez gene IDs.
adjM.m	adjacency matrix with number of rows and columns equal to length of statM.v and statR.v, ordered in same way and with same gene identifier. The resulting graph is assumed to be connected.
nseeds	number of seeds/modules to search for. This should be a number such that P-values of significance after multiple testing is less than some reasonable FDR threshold, i.e. 0.3.
gamma	tuning parameter of spin-glass algorithm. Default value generally leads to modules in the desired size range (10-100).
nMC	number of Monte Carlo runs for establishing statistical significance of modularity values under randomisation of the molecular profiles on the network.
sizeR.v	desired size range for modules
minsizeOUT	minimum size of modules to report as interesting
writeOUT	a logical to indicate whether to write out tables in text format
nameSTUDY	a name for the study.
ew.v	The adjacency edge weight vector

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
```

DoExpMod *DoExpMod*

Description

Capture the RNA expression hotspot based on the differtional expression statistics in the contex of human interactome

Usage

```
DoExpMod(statR.m, adj.m, nseeds = 100, gamma = 0.5, nMC = 1000, sizeR.v = c(1, 100), minsizeOUT = 10, wri
```

Arguments

	Arguments:
	a matrix of statistics and p-values of differential mRNA expression (same dimension as statM.m and ordered in same way) with rownames annotated with entrez gene IDs.
adj.m	adjacency matrix with number of rows and columns equal to length of statM.v and statR.v, ordered in same way and with same gene identifier. The resulting graph is assumed to be connected.
nseeds	number of seeds/modules to search for. This should be a number such that P-values of significance after multiple testing is less than some reasonable FDR threshold, i.e. 0.3.
gamma	tuning parameter of spin-glass algorithm. Default value generally leads to modules in the desired size range (10-100).
nMC	number of Monte Carlo runs for establishing statistical significance of modularity values under randomisation of the molecular profiles on the network.
sizeR.v	desired size range for modules
minsizeOUT	minimum size of modules to report as interesting
writeOUT	a logical to indicate whether to write out tables in text format
nameSTUDY	a name for the study.
ew.v	The adjacency edge weight vector

DoFEMbi *DoFEMbi*

Description

DoFEMbi identifies interactome hotspots of differential promoter methylation and differential expression, where an inverse association between methylation and gene expression is assumed.

Usage

```
DoFEMbi(statM.m, statR.m, adj.m, nseeds = 100, gamma = 0.5, nMC = 1000, sizeR.v = c(1,100), minsizeOUT =
```

Arguments

	Arguments
	a matrix of statistics and p-values of differential methylation (one row for each gene promoter) with rownames annotated with entrez gene IDs.
statR.m	a matrix of statistics and p-values of differential mRNA expression (same dimension as statM.m and ordered in same way) with rownames annotated with entrez gene IDs.
adj.m	adjacency matrix with number of rows and columns equal to length of statM.v and statR.v, ordered in same way and with same gene identifier. The resulting graph is assumed to be connected.
nseeds	number of seeds/modules to search for. This should be a number such that P-values of significance after multiple testing is less than some reasonable FDR threshold, i.e. 0.3.
gamma	tuning parameter of spin-glass algorithm. Default value generally leads to modules in the desired size range (10-100).
nMC	number of Monte Carlo runs for establishing statistical significance of modularity values under randomisation of the molecular profiles on the network.
sizeR.v	desired size range for modules
minsizeOUT	minimum size of modules to report as interesting
writeOUT	a logical to indicate whether to write out tables in text format
nameSTUDY	a name for the study.
ew.v	The adjacency edge weight vector

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.
data(toydata);
DoFEMbi(toydata$statM,toydata$statR,toydata$adjacency,nseeds=1,gamma=0.5,nMC=1000,sizeR.v=c(1,100),minsizeOUT=
```

DoIntEpi450k *DoIntEpi450k*

Description

Generate differentcial methylation statistics using 450K methylation matrix.

Usage

```
DoIntEpi450k(dnaM.m, phenoM.v, adj.m)
```

Arguments

	Arguments:
	normalised DNA methylation 450k data matrix, with rownames annotated to 450k probe IDs.
pheMoM.v	phenotype vector corresponding to dnaM.m
adj.m	adjacency matrix of a network of relations (e.g. PPI network) with rownames/colnames annotated to NCBI Entrez gene IDs. Note: The PPI network can be derived from the Pathway Commons resource*Cerami2011* and follows the procedure described in *West2013*. The PIN used in previous papers is available at http://sourceforge.net/projects/funepi The PPI network consists of 8434 genes annotated to NCBI Entrez identifiers, and is sparse containing 303600 documented interactions (edges). If the user wishes they can use a different PPI network or generate statR and statM using different method.

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
```

DoIntExp *DoIntExp*

Description

generate the statR, adjacency, annotation matrix for the DoExpMod.

Usage

```
DoIntExp(exp.m, phenoR.v, adj.m)
```

Arguments

	Arguments:
	normalized gene expression data matrix with rownames annotated to NCBI Entrez gene IDs. If the mapped Entrez gene IDs are not unique, we use the average value of the same Entrez gene ID as the expresssion value.
phenoR.v	phenotype vector corresponding to dnaM.m
adj.m	adjacency matrix of a network of relations (e.g. PPI network) with rownames/colnames annotated to NCBI Entrez gene IDs. Note: The PPI network can be derived from the Pathway Commons resource*Cerami2011* and follows the procedure described in *West2013*. The PIN used in previous papers is available at http://sourceforge.net/projects/funepi The PPI network consists of 8434 genes annotated to NCBI Entrez identifiers, and is sparse containing 303600 documented interactions (edges). If the user wishes they can use a different PPI network or generate statR and statM using different method.

DoIntFEM450k *DoIntFEM450k*

Description

generate the statM, statR, adjacency for the DoFEMbi.

Usage

```
DoIntFEM450k(dnaM.m, exp.m, phenoM.v, phenoR.v, adj.m)
```

Arguments

	Arguments:
	normalised DNA methylation 450k data matrix, with rownames annotated to 450k probe IDs.
expMm	normalized gene expression data matrix with rownames annotated to NCBI Entrez gene IDs. If the mapped Entrez gene IDs are not unique, we use the average value of the same Entrez gene ID as the expresssion value.
phenoM.v	phenotype vector corresponding to dnaM.m
phenoR.v	phenotype vector corresponding to dnaM.m
adj.m	adjacency matrix of a network of relations (e.g. PPI network) with rownames/colnames annotated to NCBI Entrez gene IDs. Note: The PPI network can be derived from the Pathway Commons resource*Cerami2011* and follows the procedure described in *West2013*. The PIN used in previous papers is available at http://sourceforge.net/projects/funepi The PPI network consists of 8434 genes annotated to NCBI Entrez identifiers, and is sparse containing 303600 documented interactions (edges). If the user wishes they can use a different PPI network or generate statR and statM using different method.

DoLimma *generate t value and p value using lmFit in Limma, this fucntion is used by DoIntFEM450k*

Description

generate t value and p value using lmFit in Limma

Usage

```
DoLimma(data.m, pheno.v)
```

Arguments

```
data.m
```

```
pheno.v
```

Examples

```
##---- Should be DIRECTLY executable !! ----
##-- ==>  Define data, use random,
##--or do  help(data=index)  for the standard data sets.

## The function is currently defined as
```

Entrez.GeneSybo.list *EntrezID and the GeneSymbol mapping list data*

Description

EntrezID and the GeneSymbol mapping list data from package org.Hs.eg.db

Usage

```
data(Entrez.GeneSybo.list)
```

Format

The format is: List of 46265 $ 1 : chr "A1BG" $ 10 : chr "NAT2" $ 100 : chr "ADA" $ 1000 : chr "CDH2" $ 10000 : chr "AKT3" $ 100008586: chr "GAGE12F" $ 100008587: chr "RNA5-8S5" $ 100008588: chr "RNA18S5" $ 100008589: chr "RNA28S5" $ 100009601: chr "TRNAY1" $ 100009602: chr "TRNAY2" $ 100009603: chr "TRNAA2" $ 100009604: chr "TRNAA3" $ 100009605: chr "TRNAF1" $ 100009606: chr "TRNAF2" $ 100009607: chr "TRNAH5" $ 100009613: chr "ANO1-AS2" $ 100009667: chr "POU5F1P5" $ 100009668: chr "POU5F1P6" $ 100009669: chr "POU5F1P7" $ 100009670: chr "POU5F1P8" $ 100009675: chr "MRT4" $ 100009676: chr "ZBTB11-AS1" $ 10001 : chr "MED6" $ 10002 : chr "NR2E3" $ 10003 : chr "NAALAD2" $ 100033391: chr "VN2R2P" $ 100033392: chr "VN2R3P" $ 100033393: chr "VN2R4P" $ 100033394: chr "VN2R5P" $ 100033395: chr "VN2R6P" $ 100033396: chr "VN2R7P" $ 100033398: chr "VN2R10P" $ 100033399: chr "VN2R11P" $ 100033400: chr "VN2R12P" $ 100033401: chr "VN2R13P" $ 100033402: chr "VN2R14P" $ 100033403: chr "VN2R15P" $ 100033404: chr "VN2R16P" $ 100033406: chr "VN2R18P" $ 100033407: chr "VN2R19P" $ 100033408: chr "VN2R20P" $ 100033409: chr "OTX2P1" $ 100033410: chr "SATB1P1" $ 100033411: chr "DUXB" $ 100033413: chr "SNORD116-1" $ 100033414: chr "SNORD116-2" $ 100033415: chr "SNORD116-3" $ 100033416: chr "SNORD116-4" $ 100033417: chr "SNORD116-5" $ 100033418: chr "SNORD116-6" $ 100033419: chr "SNORD116-7" $ 100033420: chr "SNORD116-8" $ 100033421: chr "SNORD116-9" $ 100033422: chr "SNORD116-10" $ 100033423: chr "SNORD116-11" $ 100033424: chr "SNORD116-12" $ 100033425: chr "SNORD116-13" $ 100033426: chr "SNORD116-14" $ 100033427: chr "SNORD116-15" $ 100033428: chr "SNORD116-16" $ 100033429: chr "SNORD116-17" $ 100033430: chr "SNORD116-18" $ 100033431: chr "SNORD116-20" $ 100033432: chr "SNORD116-21" $ 100033433: chr "SNORD116-22" $ 100033434: chr "SNORD116-23" $ 100033435: chr "SNORD116-24" $ 100033436: chr "SNORD116-25" $ 100033437: chr "SNORD115-2" $ 100033438: chr "SNORD116-26" $ 100033439: chr "SNORD116-27" $ 100033440: chr "SNORD115-3" $ 100033441: chr "SNORD115-4" $ 100033442: chr "SNORD115-5" $ 100033443: chr "SNORD115-6" $ 100033444: chr "SNORD115-7" $ 100033445: chr "SNORD115-8" $ 100033446: chr "SNORD115-9" $ 100033447: chr "SNORD115-10" $ 100033448: chr "SNORD115-11" $ 100033449: chr "SNORD115-12" $ 100033450: chr "SNORD115-13" $ 100033451: chr "SNORD115-14" $ 100033453: chr "SNORD115-15" $ 100033454: chr "SNORD115-16" $ 100033455: chr "SNORD115-17" $ 100033456: chr "SNORD115-18" $ 100033458: chr "SNORD115-19" $ 100033460: chr "SNORD115-20" $ 100033603: chr "SNORD115-21" $ 100033799: chr "SNORD115-22" $ 100033800: chr "SNORD115-23" $ 100033801: chr "SNORD115-25" $ 100033802: chr "SNORD115-26" $ 100033803: chr "SNORD115-29" $ 100033804: chr "SNORD115-30" $ 100033805: chr "SNORD115-31" $ 100033806: chr "SNORD115-32" [list output truncated]

Examples

```
data(Entrez.GeneSybo.list)
## maybe str(Entrez.GeneSybo.list) ; plot(Entrez.GeneSybo.list) ...
```

`fembi.o` *fembi.o*

Description

One FEM result on real cancer methylation and gene expression data

Usage

```
data(fembi.o)
```

Format

The format is:

Examples

```
data(fembi.o)
## maybe str(fembi.o) ; plot(fembi.o) ...
```

FemModShow *FemModShow*

Description

generate particular module net which is from FEM result object such as fembi.o which can be loaded by "data(fembi.o)". and also it will return an igraph object.

Usage

```
FemModShow(mod, name = "mod", edgeweight, adjacency,mode="integration")
```

Arguments

mod	particular module of the FEM result object
name	the name of the module
edgeweight	FEM result object's edgeweight
adjacency	the whole net adjacency matrix
mode	There are three mode, "integration", "Epi","Epx". "integration" means the module is from DoFEMbi, "Epi" means the module is from DoEpiMod, "Exp" means the module is from DoExpMod.

Examples

```
data(fembi.o)
data(realdata)
FemModShow(fembi.o$topmod$HAND2,name="HAND2",fembi.o$ew,realdata$adjacency)
```

map450kEID.v *map450kEID*

Description

Enrez ID and gene symbol

Usage

```
data(map450kEID.v)
```

Examples

```
data(map450kEID.v)
## maybe str(map450kEID.v) ; plot(map450kEID.v) ...
```

probeInfoALL.lv *Probes all information.*

Description

A list include the 450k methylation probes's Design, ID, and GeneGroup, etc.

Usage

```
data(probeInfoALL.lv)
```

Examples

```
data(probeInfoALL.lv)
## maybe str(probeInfoALL.lv) ; plot(probeInfoALL.lv) ...
```

realdata *realdata from TCGA endometrial cancer*

Description

realdata from TCGA endometrial cancer. Including statitics files of Methylation, RNA Expression, and also the adjacency matrix file and annotation file.

Usage

```
data(realdata)
```

Examples

```
data(realdata)
## maybe str(realdata) ; plot(realdata) ...
```

tennodes *tennodes*

Description

Randomly selected 10 nodes in toydata.

Usage

```
data(tennodes)
```

Examples

```
data(tennodes)
## maybe str(tennodes) ; plot(tennodes) ...
```

toydata *toydata*

Description

Artifical created statitics of Methylation, RNA Expression, and also the adjacency matrix and annotation matrix. Thees data are used to test and prove that FEM's ability to find hotspot or module based on inverse association between methylation and gene expression.

Usage

```
data(toydata)
```

Format

The format is: List of 4 $ statM : num [1:84, 1:2] -0.06511 0.00116 0.19583 3.93402 -0.0254 ... ..- attr(*, "dimnames")=List of 2 .. ..$: chr [1:84] "1" "2" "3" "4" ... .$: NULL $ statR : num [1:84, 1:2] -0.0959 -0.033 0.1779 -2.5759 -0.1286 ... ..- attr(*, "dimnames")=List of 2 .. ..$: chr [1:84] "1" "2" "3" "4" ... .$: NULL $ adjacency :Formal class 'dgCMatrix' [package "Matrix"] with 6 slots .. ..@ i : int [1:300] 26 30 79 5 40 3 10 18 27 29 ... .@ p : int [1:85] 0 2 3 5 17 20 21 23 25 27 ... .@ Dim : int [1:2] 84 84 .. ..@ Dimnames:List of 2 .. .. ..$: chr [1:84] "1" "2" "3" "4" ... .$: NULL .. ..@ x : num [1:300] 1 1 1 1 1 1 1 1 1 1 ... .@ factors : list() $ annotation: chr [1:84, 1:2] "1" "2" "3" "4" ... ..- attr(*, "dimnames")=List of 2 .. ..$: NULL .. ..$: chr [1:2] "EentrezID" "GeneSymbol"

Examples

```
data(toydata)
## maybe str(toydata) ; plot(toydata) ...
```

Index