Package 'FEM'

April 23, 2016

Type Package

Title Identification of Functional Epigenetic Modules

Version 2.6.0

Date 2015-08-06

Author Andrew E. Teschendorff and Yinming Jiao

Maintainer Andrew E. Teschendorff <andrew@picb.ac.cn>

Description The FEM package performs a systems-level integrative analysis of DNA methylation and gene expression data. It seeks modules of functionally related genes which exhibit differential promoter DNA methylation and differential expression, where an inverse association between promoter DNA methylation and gene expression is assumed. For full details, see Jiao et al Bioinformatics 2014.

License GPL (>=2)

Depends
AnnotationDbi,Matrix,marray,corrplot,igraph,impute,limma,org.Hs.eg.db,graph,BiocGenerics

Imports graph

biocViews SystemsBiology,NetworkEnrichment,DifferentialMethylation,DifferentialExpression

NeedsCompilation no

R topics documented:

FEM-package	*Identification of interactome hotspots of differential promoter DNA methylation and differential expression*

Description

The FEM package performs a systems-level integrative analysis of DNA methylationa and gene expression. It seeks modules of functionally related genes which exhibit differential promoter DNA methylation and differential expression, where an inverse association between promoter DNA methylation and gene expression is assumed. For full details, see Jiao et al Bioinformatics 2014.

Details

Package:	FEM
Type:	Package
Version:	2.1.4
Date:	2015-03-27
License:	GPL >2

Author(s)

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

References

1) Jiao 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 Aug 15;30(16):2360-2366.

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	*Identifies differential DNA methylation network hotspots.*

Description

This function aims to identify subnetworks where many members exhibit differential DNA methylation in relation to the phenotype of interest.

Usage

```
DoEpiMod(intEpi.o, nseeds = 100, gamma = 0.5, nMC = 1000,
 sizeR.v = c(1, 100), minsizeOUT = 10, writeOUT = TRUE, nameSTUDY = "X", ew.v = NULL)
```

Arguments

intEpi.o	The output of the DoIntEpi450k function
nseeds	An integer specifying the number of seeds and therefore modules to search for. By default this number is 100.
gamma	A parameter of the spin-glass algorithm, which determines the average module size. Default value generally leads to modules in the desired size range (10-100 genes).
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, to be used as label in the output files.
ew.v	The edge weight vector of the integrated network. This is actually generated by the function itself, and can speed up inference significantly, if provided as argument to a 2nd instance of the function. Default value is NULL.

Value

A list with following entries:

size	A vector of inferred module sizes for each of the ntop seeds.
mod	A vector of associated modularities.
pv	A vector of associated significance P-values with resolution limited by nMC runs.
selmod	Index positions of significant modules of size at least minsizeOUT
fem	A summary matrix of the selected modules.
topmod	A list of summary matrices for each of the selected modules.
sgc	A list of the spin-glass module detection algorithm for each seed.

| ew | The edge-weight vector of the integrated network. |
| adj | adjacency matrix of the maximally connected integrated network (at present only maximally connected subnetwork is used).It is same to intEpi.o$adj, and wil be used for FemModShow function |

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

Examples

```
data(Toydata);
intEpi.o <- list(statM=Toydata$statM,adj=Toydata$adjacency);
EpiMod.o=DoEpiMod(intEpi.o,nseeds=1,gamma=0.5,nMC=1000,sizeR.v=c(1,100),
 minsizeOUT=10,writeOUT=TRUE,nameSTUDY="TEST",ew.v=NULL);

#You can also test on the Realdata which contains DNA methylation of 17
#normal and 118 endometrial cancer samples. Since running on the realdata is time-consuming,  we comment it out.
#data(Realdata);
# intEpi.o <- list(statM=Realdata$statM,adj=Realdata$adjacency);
#EpiMod.o=DoEpiMod(intEpi.o,nseeds=100,gamma=0.5,nMC=1000,sizeR.v=c(1,100),nameSTUDY="TEST")
```

DoExpMod   *Identifies differential mRNA expression network hotspots.*

Description

This function aims to identify subnetworks where many members exhibit differential mRNA expression in relation to the phenotype of interest.

Usage

```
DoExpMod(intExp.o, nseeds = 100, gamma = 0.5, nMC = 1000, sizeR.v = c(1,100),
  minsizeOUT = 10, writeOUT = TRUE, nameSTUDY = "X", ew.v = NULL)
```

Arguments

| intExp.o | The output of the DoIntExp function |
| nseeds | An integer specifying the number of seeds and therefore modules to search for. By default this number is 100. |

gamma	A parameter of the spin-glass algorithm, which determines the average module size. Default value generally leads to modules in the desired size range (10-100 genes).
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, to be used as label in the output files.
ew.v	The edge weight vector of the integrated network. This is actually generated by the function itself, and can speed up inference significantly, if provided as argument to a 2nd instance of the function. Default value is NULL.

Value

A list with following entries:

size	A vector of inferred module sizes for each of the ntop seeds.
mod	A vector of associated modularities.
pv	A vector of associated significance P-values with resolution of nMC
selmod	Index positions of significant modules of size at least minsizeOUT
nd	smaller than the maximum specified in sizeR.v
fem	A summary matrix of the selected modules.
topmod	A list of summary matrices for each of the selected module
sgc	A list of the spin-glass module detection algorithm for each seed.
ew	The edge-weight vector of the integrated network.
adj	adjacency matrix of the maximally connected integrated network (at present only maximally connected subnetwork is used).It is same to intFEM.o$adj, and wil be used for FemModShow function

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

Examples

```
data(Toydata);
intExp.o <- list(statR=Toydata$statR,adj=Toydata$adj);
ExpMod.o=DoExpMod(intExp.o,nseeds=1,gamma=0.5,nMC=1000,sizeR.v=c(1,100),
 minsizeOUT=10,writeOUT=TRUE,nameSTUDY="TEST",ew.v=NULL);
#You can also test on the Realdata which contains RNA expression of 17
#normal and 118 endometrial cancer samples. Since running on the realdata is time-consuming,  we comment it out.
#data(Realdata);
#intExp.o <- list(statM=Realdata$statR,adj=Realdata$adjacency);
#EpiMod.o=DoEpiMod(intExp.o,nseeds=100,gamma=0.5,nMC=1000,sizeR.v=c(1,100),nameSTUDY="TEST")
```

DoFEMbi   *Identifies interactome hotspots of differential promoter DNAm methy-*
                                        lation and differential mRNA expression.

Description

DoFEMbi identifies interactome hotspots of differential promoter methylation and differential ex-
pression in relation to a phenotype of interest, where an inverse association between methylation
and gene expression is assumed.

Usage

```
DoFEMbi(intFEM.o, nseeds = 100, gamma = 0.5, nMC = 1000, sizeR.v = c(1,100),
  minsizeOUT = 10, writeOUT = TRUE, nameSTUDY = "X", ew.v = NULL)
```

Arguments

intFEM.o	The output of the DoFEMbi function.
nseeds	An integer specifying the number of seeds and therefore modules to search for. By default this number is 100.
gamma	A parameter of the spin-glass algorithm, which determines the average module size. Default value generally leads to modules in the desired size range (10-100 genes).
nMC	Number of Monte Carlo runs for establishing statistical significance of modu-larity 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, to be used as label in the output files.
ew.v	The edge weight vector of the integrated network. This is actually generated by the function itself, and can speed up inference significantly, if provided as argument to a 2nd instance of the function. Default value is NULL.

Value

A list with following entries:

size	A vector of inferred module sizes for each of the ntop seeds.
mod	A vector of associated modularities.
pv	A vector of associated significance P-values with resolution of nMC
selmod	Index positions of significant modules of size at least minsizeOUT
fem	A summary matrix of the selected modules.
topmod	A list of summary matrices for each of the selected modules.
sgc	A list of the spin-glass module detection algorithm for each seed.
ew	The edge-weight vector of the integrated network.
adj	adjacency matrix of the maximally connected integrated network (at present only maximally connected subnetwork is used).It is same to intFEM.o$adj, and wil be used for FemModShow function

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

Examples

```
data(Toydata);
intFEM.o <- list(statM=Toydata$statM,statR=Toydata$statR,adj=Toydata$adj);
DoFEMbi(intFEM.o,nseeds=1,gamma=0.5,nMC=1000,sizeR.v=c(1,100),
 minsizeOUT=10,writeOUT=TRUE,nameSTUDY="TEST",ew.v=NULL);
#You can also test on the Realdata contains matched DNA methylation and RNA Expression of 17 normal and 118 cancer
#data(Realdata);
#intFEM.o <- list(statM=Realdata$statM,statR=Realdata$statR,adj=Realdata$adjacency);
#DoFEMbi(intFEM.o,nseeds=100,gamma=0.5,nMC=1000,sizeR.v=c(1,100),nameSTUDY="TEST");
```

DoIntEpi450k	*Integrates statistics of differential DNA methylation for a specified contrast with a network adjacency matrix.*

Description

Integrates statistics of differential DNA methylation for a specified contrast with a network adjacency matrix. Output of this function is required input for DoEpiMod.

Usage

```
DoIntEpi450k(statM.o,adj.m,c)
```

Arguments

statM.o	The output of GenStatM, or analogous object generated by user.
adj.m	The adjacency matrix defining the network of gene relations, e.g. a PPI network, annotated to NCBI entrez gene IDs.
c	An integer specifying the desired contrast, i.e. column of the contrast matrix. User needs to check which column of statM.o$cont is the one of interest.

Value

A list with following entries:

statM	matrix of DNA methylation moderated t-statistics and P-values for the genes in the integrated network
adj	adjacency matrix of the maximally connected integrated network (at present only the maximally connected subnetwork is used in the subsequent inference
avbeta	average DNA methylation data matrix mapped to unique Entrez IDs

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

DoIntExp	*Integrates statistics of differential expression for a specified contrast with a network adjacency matrix.*

Description

Integrates statistics of differential expression for a specified contrast with a network adjacency matrix. Output of this function is required input for DoExpMod.

Usage

```
DoIntExp(statR.o,adj.m,c)
```

Arguments

statR.o	The output of GenStatR, or analogous object generated by user.
adj.m	The adjacency matrix defining the network of gene relations, e.g. a PPI network, annotated to NCBI entrez gene IDs.
c	An integer specifying the desired contrast, i.e. column of the contrast matrix. User needs to check which column of statR.o$cont is the one of interest.

Value

A list with following entries:

statR	matrix of gene expression moderated t-statistics and P-values for the genes in the integrated network
adj	adjacency matrix of the maximally connected integrated network (at present only the maximally connected subnetwork is used for subsequent inference)
avexp	average expression data matrix mapped to unique Entrez IDs

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366.

DoIntFEM450k	*Integrates statistics of differential DNA methylation and differential expression for a specified contrast with a network adjacency matrix.*

Description

Integrates statistics of differential DNA methylation and differential expression for a specified contrast with a network adjacency matrix. Output of this function is required input for DoFEMbi.

Usage

```
DoIntFEM450k(statM.o,statR.o,adj.m,cM,cR)
```

Arguments

statM.o The output of GenStatM, or analogous object generated by user.

statR.o The output of GenStatR, or analogous object generated by user.

adj.m The adjacency matrix defining the network of gene relations, e.g. a PPI network, annotated to NCBI entrez gene IDs.

cR An integer specifying the desired contrast, i.e. column of the contrast matrix. User needs to check which column of statR.o$cont is the one of interest.

cM An integer specifying the desired contrast, i.e. column of the contrast matrix. User needs to check which column of statM.o$cont is the one of interest. Note that for this to make sense the contrast cM must correspond to the same one as specified by cR.

Value

A list with following entries:

statM matrix of DNA methylation moderated t-statistics and P-values for the genes in the integrated network

statR matrix of gene expression moderated t-statistics and P-values for the genes in the integrated network

adj adjacency matrix of the maximally connected integrated network (at present only maximally connected subnetwork is used).

avexp average expression data matrix mapped to unique Entrez IDs

avbeta average DNA methylation data matrix mapped to unique Entrez IDs

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

FemModShow *Generates a figure showing the inferred FEM module(s)*

Description

This function generates a network representation of an inferred FEM module.

Usage

```
FemModShow(mod, name, fem.o, mode="Integration")
```

Arguments

mod	the module of the FEM result object for which a figure is to be generated
name	the name for the output figure file
fem.o	FEM result object generated by DoFEMbi, DoEpiMod, or DoExpMod
mode	There are three modes: "Integration","Epi","Epx". "Integration" means the module is from DoFEMbi, "Epi" means the module is from DoEpiMod, "Exp" means the module is from DoExpMod.

Value

An igraph object representing the module. At the same time, the module graph is generated and saved as a "X.pdf" file with "X" the name given in the argument to the function.

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

Examples

```
data(Realdata)
FemModShow(Realdata$fembi.o$topmod$HAND2,name="HAND2",Realdata$fembi.o)
```

GenStatM   *Generation of differential DNA methylation statistics*

Description

Given an Illumina 450k data matrix and a phenotype vector, it will generate statistics of differential DNA methylation using limma, at the gene-level, for all pairwise comparisons of phenotype levels.

Usage

```
GenStatM(dnaM.m,pheno.v)
```

Arguments

dnaM.m	The DNA methylation beta valued data matrix with rownames annotated to Illumina 450k probe IDs
pheno.v	The phenotype vector.

Value

top	A list of matrices, from the limma output, ranking genes according to differential methylation between two phenotypes as specified in the contrasts matrix, i.e. top[[i]] will contain the ranking for the i'th contrast.
cont	The contrasts matrix, with columns labeling the contrasts.
avbeta	The beta-valued DNAm data matrix at the gene-level, following the procedure as described in Jiao et al Bioinformatics 2014.

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

GenStatR *Generation of differential expression statistics*

Description

Given a gene expression data matrix annotated to entrez gene IDs and a phenotype vector, it will generate statistics of differential expression using limma, at the gene-level, for all pairwise comparisons of phenotype levels.

Usage

```
GenStatR(exp.m,pheno.v)
```

Arguments

exp.m	The gene expression data matrix with rownames annotated to entrez gene IDs. If there are multiple rows with the same gene ID, function will average these.
pheno.v	The phenotype vector.

Value

top	A list of matrices, from the limma output, ranking genes according to differential expression between two phenotypes as specified in the contrasts matrix, i.e. top[[i]] will contain the ranking for the i'th contrast.
cont	The contrasts matrix, with columns labeling the contrasts.
avexp	The averaged expression matrix at the gene-level.

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

---

probe450kfemanno	*Processed annotation of Illumina 450k probes for use in the FEM algorithm.*

Description

A list of entries giving various annotation information of the 450k probes

Usage

```
data(probe450kfemanno)
```

Value

typeC	a vector of 0's and 1's indicating whether a probe maps to a CpG (1) or not (0).
CGI	a vector of 0's and 1's indicating whether a probe maps to a CGI (1) or not (0).
design	a vector of 1's and 2's, specifying design type of probe (1 = type I, 2 = type II).
probeID	the 450k probe IDs.
GeneGroup	an integer vector specifying which gene region the probe maps to. Regions are annotated as follows: 1=TSS1500, 2=TSS200, 3=5'UTR, 4=1stExon, 5=gene body, 6=3'UTR. Probes with ambiguous mappings are assigned an NA.
eid	mapping of 450k probes to Entrez Gene ID.

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

Realdata   *A dataset object derived from real DNA methylation and gene expression data from normal endometrial and endometrial cancer tissue*

Description

Realdata is a list containing data derived from mRNA Expression and DNA methylation from an endometrial cancer study. See Jones et al PLoS Med.2013 for details about the data.

Usage

```
data(Realdata)
```

Value

statM   a matrix of statistics and p-values of differential DNA Methylation between endometrial cancer and normal tissue (cancer compared to normal) with rownames annotated with entrez gene IDs.

statR   a matrix of statistics and p-values of differential RNA Expression between endometrial cancer and normal tissue (cancer compared to normal) with rownames annotated with entrez gene IDs.

adjacency   adjacency matrix of PPI network, with number of rows equal to the number of rows of Realdata$statM, ordered in same way and with same gene identifier. The resulting graph is connected.

fembi.o   this entry represents the output of the function DoFEMbi() (see example below). This is included so as to avoid having to rerun the (lengthy) example from scratch.

References

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.

Examples

```
#data(Realdata);
#intFEM.o <- list(statM=Realdata$statM,statR=Realdata$statR,adj=Realdata$adjacency);
```

Toydata	*Artificial data set.*

Description

A list object containing artificial data to test the FEM algorithm. It contains artifically created statistics of differential DNA methylation, mRNA expression, as well as an artificial adjacency matrix to illustrate the application of FEM.

Usage

```
data(Toydata)
```

Value

statM	a matrix of simulated statistics and p-values of differential DNA Methylation with 84 rows (genes).
statR	a matrix of simulated statistics and p-values of differential RNA Expression with 84 rows (genes).
adjacency	artificial adjacency matrix of 84 rows and columns, ordered according to the rows of previous two matrices. Adjacency matrix is a random graph except for containing a clique defined by 10 nodes.
tennodes	a vector of 10 integer indices, labeling the nodes/genes which form the clique in the network.

Author(s)

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

References

A systems-level integrative framework for genome-wide DNA methylation and gene expression data identifies differential gene expression modules under epigenetic control. Jiao Y, Widschwendter M, Teschendorff AE. Bioinformatics. 2014;30(16):2360-2366

Examples

```
data(Toydata);
intFEM.o <- list(statM=Toydata$statM,statR=Toydata$statR,adj=Toydata$adjacency);
DoFEMbi(intFEM.o,nseeds=1,gamma=0.5,nMC=1000,sizeR.v=c(1,100),minsizeOUT=10,
 writeOUT=TRUE,nameSTUDY="TEST",ew.v=NULL);
```

Index