Package 'NEpiC'

January 20, 2025

Type Package

Title Network Assisted Algorithm for Epigenetic Studies Using Mean and
Variance Combined Signals

Version 1.0.1

Date 2016-03-20

Author Peifeng Ruan

Depends igraph, PairedData

Maintainer Peifeng Ruan <pruan12@fudan.edu.cn>

Description Package for a Network assisted algorithm for Epigenetic studies using mean and vari-
ance Combined signals: NEpiC. NEpiC combines both signals in mean and variance differ-
ences in methylation level between case and control groups searching for differentially methy-
lated sub-networks (modules) using the protein-protein interaction network.

License GPL-2

NeedsCompilation no

Repository CRAN

Date/Publication 2016-03-26 15:13:10

Contents

---

candidate_genes   *Chosing candidate genes.*

Description

Chosing genes in top ranked modules as candidate genes for outcomes. Further prioritized candidate
genes can be chosen based on percentages selected by top ranked modules selected if prioritizing =
T.

Usage

```
candidate_genes(res, pct = 0.01, prioritizing)
```

Arguments

res   The result for c_dms.

pct   A float which determines the number of candidate genes. For example, if pct is
                0.01 as default setting, the genes in top 1 percent modules will be chosen as the
                candidate genes.

prioritizing   To further prioritize the candidate genes or not.

Value

A list for candidate genes or prioritized candidate genes.

References

NEpiC: a Network-assisted algorithm for Epigenetic studies using mean and variance Combined
signals Peifeng Ruan, Jing Shen, Regina M. Santella, Shuigeng Zhou, Shuang Wang.(2015)

Examples

```
# candidate_genes(res, pct = 0.01, prioritizing = T)
```

---

c_dms   *Combined signals dense modules searching algorithm.*

Description

c_dms uses the combined signals in both mean and variance differences in methylation measures
between case and control groups to search for differentially methylated sub-networks (modules)
using a protein-protein interaction (PPI) network. It returns ordered dense modules.

Usage

```
c_dms(network, gene2weight, d = 1, r = 0.1)
```

Arguments

network	A dataframe describing the edges of s PPI network. Two columns are needed in this dataset: interactor A and interactor B.
gene2weight	A dataframe describing combined gene-level scores. The first column of this dataframe, named 'gene', contains the gene names, which should use the same annotation as those used in the network. The second column 'weight' contains gene-level combined scores.
d	An integer determining the order of neighbour genes to be searched in NEpiC algorithm. We suggest to set d=1 for NEpiC.
r	A number determining the threshold for including a new gene into the module.

Value

A list containing several statistics for modules, including genes in modules, module scores before and after normalization.

References

Peifeng Ruan, Jing Shen, Regina M. Santella, Shuigeng Zhou, Shuang Wang.(2015). NEpiC: a Network-assisted algorithm for Epigenetic studies using mean and variance Combined signals.
Peilin Jia, Siyuan Zheng, Jirong Rong, Wei Zheng, Zhongming Zhao. (2011) Bioformatics. dmG-WAS: dense module searching for genome-wide association studies in protein-protein interaction networks.

Examples

```
#res=c_dms(network, gene2weight, d = 1, r = 0.1)
```

---

c_score	*Computing the mean and variance combined signal scores*

Description

A function computing gene-level scores combining both signals in mean and variance differences.

Usage

```
c_score(mean_sig, var_sig, lambda = "default")
```

Arguments

mean_sig	A dataframe describing mean signals for every CpG sites. Three columns are needed in this dataset, which are 'cpg', 'gene' and 'mean_p'. 'cpg' column has names of CpG sites. 'gene' column has has gene annotations for CpG sites. 'mean_p' column has p values of mean differences for CpG sites.

var_sig   A dataframe describing variance signals for every CpG sites. Three columns
                 are needed in this dataset, which are 'cpg', 'gene' and 'var_p'. 'cpg' column
                 has names of CpG sites. 'gene' column has has gene annotations for CpG sites.
                 'var_p' column has p values of variance differences for CpG sites.

lambda   A parameter to balance mean and variance scores. NEpiC computes it as default
                 if it is not otherwise specified by user. If lambda = 1, NEpiC will search dense
                 modules using only mean signals. If lambda = 0, NEpiC will search dense
                 modules using only variance signals. Users can also specify any float between 0
                 and 1 for lambda if they have any additional prior biology information.

Value

c_score returns a dataframe describing the combined gene-level scores, which can be used directly
in c_dms.

References

NEpiC: a Network-assisted algorithm for Epigenetic studies using mean and variance Combined
signals. Peifeng Ruan, Jing Shen, Regina M. Santella, Shuigeng Zhou, Shuang Wang.(2015)

Examples

```
#c_score(mean_sig, var_sig, lambda = "default")
```

---

permutation   *Validating identified modules using permutations*

Description

A function shuffling the tumor/adjacent normal labels within a tumor/adjacent normal pair and
repeating the permutation procedures to calculate the permutated module scores for the identified
top 1 percentage modules.

Usage

```
permutation(normal_data, tumor_data, annotation, res, n = 100)
```

Arguments

normal_data   The Illumina 450k methylation array results for normal samples. The rows rep-
                 resent the methylation sites and the columns represent different samples.

tumor_data   The Illumina 450k methylation array results for adjacent tumor paire samples.
                 The rows represent the methylation sites and the columns represent different
                 samples.

annotation   The annotation file for Illumina 450k methylation array data. The first column
                 is the methylation sites' names, whose column name is "cpg" and the second
                 column is the methylation sites' gene annotations, whose column name is "gene"

res   The result of the c_dms, which contains the information of identified modules.

n   The numbers of permutations.

Value

A dataset contains identified modules' names, identified modules' score, identified modules' permutated scores and the last column contains the permutation p values for identified modules.

References

Peifeng Ruan, Jing Shen, Regina M. Santella, Shuigeng Zhou, Shuang Wang.(2015). NEpiC: a Network-assisted algorithm for Epigenetic studies using mean and variance Combined signals.

Examples

```
#perm_result = function (normal_data, tumor_data, annotation, res, n = 100)
```

Index