Package 'motif4node'

January 23, 2023

Type Package

Title Motif4node, motif analysis using all non-redundant four-node gene circuits

Version 0.1.3

Author Benjamin Clauss, Mingyang Lu

Maintainer Benjamin Clauss <b.clauss@northeastern.edu>

Description R package for circuit motif analysis.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

NeedsCompilation no

VignetteBuilder knitr

Imports igraph, visNetwork, sRACIPE, scales, ggplot2, cowplot, pheatmap, htmlwidgets, SummarizedExperiment, webshot, combinat, RColorBrewer, grDevices

Suggests knitr

RoxygenNote 7.2.3

Depends R (>= 3.5.0)

R topics documented:

| all.circuits | *Data file containing the topology of all 60212 non-redundant four-node gene circuits* |

Description

This data contains the topology of all 60212 non-redundant four-node gene circuits.

Usage

```
all.circuits
```

Format

A list of 60212 circuits, each element is a matrix of adjacency matrix. Four columns/rows represent genes "A", "B", "C", and "D". In each adjacency matrix, 1 represents activation, 2 represents inhibition, and 0 represents no interaction.

---

analysis_circuit_2node	
	Analysis script to evaluate the state distribution of a two-node gene circuit

---

Description

Analysis script to evaluate the state distribution of a two-node gene circuit

Usage

```
analysis_circuit_2node(rSet, numModels = 10000, filename)
```

Arguments

rSet	sRACIPE object. RACIPE simulation data.
numModels	Numeric. Number of models to be simulated. Default: 10000
filename	Character. Filename. Default: "myplot".

---

analysis_circuit_4node

> *Analysis script to evaluate the state distribution of a four-node gene circuit*

Description

Analysis script to evaluate the state distribution of a four-node gene circuit

Usage

```
analysis_circuit_4node(rSet, numModels = 10000, filename = "myplot")
```

Arguments

rSet	sRACIPE object. RACIPE simulation data.
numModels	Numeric. Number of models to be simulated. Default: 10000
filename	Character. Filename. Default: "myplot".

---

circuit_grouping   *Grouping two-node circuit motifs by their types*

Description

Grouping two-node circuit motifs by their types

Usage

```
circuit_grouping()
```

Value

Vector of factor: types for all two-node motifs.

---

dist_ks   *Calculate th KS distance of two gene expression distributions*

Description

Calculate th KS distance of two gene expression distributions

Usage

```
dist_ks(query, reference, experimental)
```

Arguments

query	sRACIPE object or PCA matrix of the query data
reference	sRACIPE object of the reference data
experimental	Logical. T: query is the PCA matrix from an experimental dataset. F: query is an sRACIPE object

Value

the distance between the gene expression distributions from the query and reference.

---

enrichment_coupling	*Motif enrichment analysis for the coupling of two two-node circuit motifs*

Description

Motif enrichment analysis for the coupling of two two-node circuit motifs

Usage

```
enrichment_coupling(
  all.circuits = all.circuits,
  all.scores,
  motif_list,
  new_ind,
  overlap_list,
  decreasing = T,
  if_overlap = 2,
  topCircuits = 600,
  nhill = 20
)
```

Arguments

all.circuits	List of the topologies of all 60212 non-redundant 4-node circuits. Default "all.circuits" from the package data.
all.scores	Data frame. 1st column: scores; 2nd column: circuit index. The data frame is ordered by the scores.
motif_list	List of 2 by 2 integer matrix. 2-node motif info generated by the function: generate_motif_list.
new_ind	Vector of integer. mapping to non-redundant 2-node motifs generated by the function: generate_index_conversion.
overlap_list	Matrix (6 by 6 integers). Info of overlapping from the motif location in the 4-node circuit generated by the function: generate_overlap_data.
decreasing	Logic. Whether circuits are ranked by the scores in a decreasing order (T) or not (F). Default T.
if_overlap	Whether consider two motifs with overlapping (1), without overlapping (0), or both (2). Default 2.
topCircuits	Integer. Number of top circuits for the enrichment analysis. Default 600.
nhill	Integer/Numeric. Hill coefficient for the Hill function as the weighting factor for motif counts, large n makes the Hill function more binary. Default 20.

Value

A data frame containing the enrichment scores of the coupling between 2-node circuit motifs (39 by 39)

---

enrichment_coupling_all_cases

        A convenient function to perform different motif coupling analyses altogether

---

Description

A convenient function to perform different motif coupling analyses altogether

Usage

```
enrichment_coupling_all_cases(
  all.circuits,
  all.scores,
  motif_list,
  new_ind,
  overlap_list,
  decreasing = decreasing,
  topCircuits = 600,
  nhill = 20
)
```

Arguments

all.circuits	List of the topologies of all 60212 non-redundant 4-node circuits. Default "all.circuits" from the package data.
all.scores	Data frame. 1st column: scores; 2nd column: circuit index. The data frame is ordered by the scores.
motif_list	List of 2 by 2 integer matrix. 2-node motif info generated by the function: generate_motif_list.
new_ind	Vector of integer. mapping to non-redundant 2-node motifs generated by the function: generate_index_conversion.
overlap_list	Matrix (6 by 6 integers). Info of overlapping from the motif location in the 4-node circuit generated by the function: generate_overlap_data.
decreasing	Logic. Whether circuits are ranked by the scores in a decreasing order (T) or not (F). Default T.
topCircuits	Integer. Number of top circuits for the enrichment analysis. Default 600.
nhill	Integer/Numeric. Hill coefficient for the Hill function as the weighting factor for motif counts, large n makes the Hill function more binary. Default 20.

Value

List containing the enrichment of coupling between 2-node motifs for all cases (without overlapping, overlapping, and both)

---

enrichment_single   *Motif enrichment analysis for single two-node circuit motifs*

Description

Motif enrichment analysis for single two-node circuit motifs

Usage

```
enrichment_single(
  all.circuits = all.circuits,
  all.scores,
  motif_list,
  new_ind,
  decreasing = T,
  topCircuits = 600,
  nhill = 20
)
```

Arguments

`all.circuits`	List of the topologies of all 60212 non-redundant 4-node circuits. Default "all.circuits" from the package data.
`all.scores`	Data frame. 1st column: scores; 2nd column: circuit index. The data frame is ordered by the scores.
`motif_list`	List of 2 by 2 integer matrix. 2-node motif info generated by the function: generate_motif_list.
`new_ind`	Vector of integer. mapping to non-redundant 2-node motifs generated by the function: generate_index_conversion.
`decreasing`	Logic. Whether circuits are ranked by the scores in a decreasing order (T) or not (F). Default T.
`topCircuits`	Integer. Number of top circuits for the enrichment analysis. Default 600.
`nhill`	Integer/Numeric. Hill coefficient for the Hill function as the weighting factor for motif counts, large n makes the Hill function more binary. Default 20.

Value

Data frame containing the enrichment scores of each 2-node circuit motif (39 by 1)

---

generate_index_conversion

Convert redundant indices to the indices for all 39 non-redundant mo-
tifs

Description

Convert redundant indices to the indices for all 39 non-redundant motifs

Usage

```
generate_index_conversion()
```

---

gen_network_scalefree	*Generate a random scale-free gene network consisting the two-node motifs of choice*

Description

Generate a random scale-free gene network consisting the two-node motifs of choice

Usage

```
gen_network_scalefree(num_nodes, motif_list, motif_choice)
```

Arguments

num_nodes	Integer. Number of nodes of the gene network to be generated.
motif_list	List of 2 by 2 integer matrix. 2-node motif info generated by the function: generate_motif_list.
motif_choice	Vector of integer. A vector of indices of the selected two-node circuit motifs

Value

Matrix. Adjacency matrix of the generated network

---

lin_score	*The scoring function for ranking circuits with a linear state distribution*

Description

The scoring function for ranking circuits with a linear state distribution

Usage

```
lin_score(rset)
```

Arguments

rset	The sRACIPE object of the simulated circuit.

Value

return(min(score_vector)): returns the score for the linear state distribution.

| motif_analysis | *A combined motif enrichment analysis for single two-node circuit motifs and motif coupling* |

Description

A combined motif enrichment analysis for single two-node circuit motifs and motif coupling

Usage

```
motif_analysis(
  all.circuits = all.circuits,
  all.scores,
  ylim = NULL,
  color_breaks = NULL,
  filename = NULL,
  decreasing = T,
  topCircuits = 600
)
```

Arguments

`all.circuits`	List of the topologies of all 60212 non-redundant 4-node circuits. Default "all.circuits" from the package data.
`all.scores`	Data frame. 1st column: scores; 2nd column: circuit index. The data frame is ordered by the scores.
`ylim`	Vector of numerics (2). Y axis limit for single motif enrichment. Default: NULL.
`color_breaks`	Vector that defines color scaling for pheatmap. Default: NULL
`filename`	Character. Prefix of filenames for plotting. Default: NULL. If provided, plots are also saved to files.
`decreasing`	Logical. Whether circuits are ranked by the scores in a decreasing order (T) or not (F). Default T.
`topCircuits`	Integer. Number of top circuits for the enrichment analysis. Default 600.

Value

List of plotting objects for single motif and motif coupling enrichment analyses.

| plot_adj | *Plot a network from the adjacency matrix* |

Description

Plot a network from the adjacency matrix

Usage

```
plot_adj(adj)
```

Arguments

adj   Matrix (4 by 4, integer). Adjacency matrix of a four node circuit.

Value

empty

---

plot_motif	*Plot a specific circuit motif*

Description

Plot a specific circuit motif

Usage

```
plot_motif(number, motif_list)
```

Arguments

number   Integer. Index of a two-node circuit motif

motif_list   List of 2 by 2 integer matrix. 2-node motif info generated by the function: generate_motif_list.

Value

empty

---

plot_net	*Network plotting function*

Description

Network plotting function

Usage

```
plot_net(tf_links = tf_links)
```

Arguments

tf_links   Data frame of circuit edge list. Three columns: Source, Target, Interaction Types – 1: Activation; 2: Inhibition

Value

empty

---

plot_RACIPE   *Scatterplots of PCA and Gene expression from any 4-node circuit*

Description

Scatterplots of PCA and Gene expression from any 4-node circuit

Usage

```
plot_RACIPE(rset)
```

Arguments

rset   The sRACIPE object of the simulated circuit

Value

list(g,p)

---

simu_rnorm   *RACIPE simulations with random kinetic parameters from Gaussian distributions*

Description

RACIPE simulations with random kinetic parameters from Gaussian distributions

Usage

```
simu_rnorm(rset, numModels = 10000)
```

Arguments

rset   sRACIPE object. sRACIPE output from a standard RACIPE simulation

numModels   Numeric. Number of models to be simulated. Default: 10000.

---

sim_4node *Simulate one of the 4-node circuits from the 60212 unique circuits*

Description

Simulate one of the 4-node circuits from the 60212 unique circuits

Usage

```
sim_4node(index, Gaussian = F, numModels = 10000, all.circuits = all.circuits)
```

Arguments

index	Numeric index number of 4-node circuit to be simulate. Takes values from 1:60212
Gaussian	Logical. If T, kinetic parameters will be sampled from a gaussian distribution. If F, kinetic parameters will be sampled from a uniform distribution
numModels	Numeric. Number of models to be simulated. Default: 10000
all.circuits	List. The topology of all 60212 circuit motifs. Default "all.circuits" from the package data.

Value

rset: sRACIPE object. RACIPE simulation results for a circuit

---

single_motif_permute *Generate permutations for p-value calculations*

Description

Generate permutations for p-value calculations

Usage

```
single_motif_permute(
  all.circuits = all.circuits,
  all.scores,
  decreasing = T,
  topCircuits = 600,
  no_perm
)
```

Arguments

`all.circuits`	List of the topologies of all 60212 non-redundant 4-node circuits. Default "all.circuits" from the package data.
`all.scores`	Data frame. 1st column: scores; 2nd column: circuit index. The data frame is ordered by the scores.
`decreasing`	Logical. Whether circuits are ranked by the scores in a decreasing order (T) or not (F). Default T.
`topCircuits`	Integer. Number of top circuits for the enrichment analysis. Default 600.
`no_perm`	Integer. Number of permutations.

Value

List of permuted enrichment scores.

`trig_score`	*The scoring function for ranking circuits with a triangular state distribution*

Description

The scoring function for ranking circuits with a triangular state distribution

Usage

```
trig_score(rset)
```

Arguments

`rset`	The sRACIPE object of the simulated circuit.

Value

return(min(score_vector)): returns the score for the triangular state distribution.

`z_score`	*Calculate z-score for a specific distance*

Description

Calculate z-score for a specific distance

Usage

```
z_score(distances, score)
```

Arguments

`distances`	list of distances
`score`	Numeric. score to calculate the z-score

Value

z-score

Index