Package 'corrDNA'

October 12, 2022

Type Package

Title Finding Associations in Position-Wise Aligned DNA Sequence Dataset

Version 1.0.1

Date 2018-03-08

Author Prabina Kumar Meher & A. R. Rao

Maintainer Prabina Kumar Meher <meherprabin@yahoo.com>

Depends R(>= 3.3.0)

Imports mvtnorm

LazyData TRUE

Description Can be useful for finding associations among different positions in a position-wise aligned sequence dataset. The approach adopted for finding associations among positions is based on the latent multivariate normal distribution.

License GPL (>= 2)

NeedsCompilation no

Repository CRAN

Date/Publication 2018-04-05 15:31:17 UTC

R topics documented:

assoc_comb *Complete association matrix.*

Description

All the six possible association matrices can be merged in to a single matrix to visualize the overall association among positions as well as among the occurences of nucleotides of different positions, in a position-wise aligned sequence dataset.

Usage

```
assoc_comb(x, rZiZj, rZiZjR, rZiZjY, rZiRZjR, rZiRZjY, rZiYZjY)
```

Arguments

x	A dataframe of position wise aligned sequence dataset having A, T, G and C only.
rZiZj	An object generated by using the function assoc_Zi.Zj.
rZiZjR	An object generated by using the function assoc_Zi.ZjR.
rZiZjY	An object generated by using the function assoc_Zi.ZjY.
rZiRZjR	An object generated by using the function assoc_ZiR.ZjR.
rZiRZjY	An object generated by using the function assoc_ZiR.ZjY.
rZiYZjY	An object generated by using the function assoc_ZiY.ZjY.

Details

All the six association matrices are required to be generated prior to merging them into a single matrix.

Value

A numeric matrix of order *3L* by *3L* for the dataset of *L* nucleotides long sequences.

Author(s)

Prabina Kumar Meher & A. R. Rao

Examples

```
data(don_dat)
kk <- don_dat[1:300,]
zizj <- assoc_Zi.Zj(x=kk)
zizjr <- assoc_Zi.ZjR(x=kk, rZiZj=zizj)
zizjy <- assoc_Zi.ZjY(x=kk, rZiZj=zizj)
zirzjr <- assoc_ZiR.ZjR(x=kk,rZiZj=zizj,rZiZjR=zizjr)
zirzjy <- assoc_ZiR.ZjY(x=kk,rZiZj=zizj,rZiZjR=zizjr,rZiZjY=zizjy)
```

```
ziyzjy <- assoc_ZiY.ZjY(x=kk,rZiZj=zizj,rZiZjY=zizjy)
fin_corr <- assoc_comb(x=kk, rZiZj=zizj,rZiZjR=zizjr,rZiZjY=zizjy,
rZiRZjR=zirzjr,rZiRZjY=zirzjy,rZiYZjY=ziyzjy)
fin_corr
```

---

assoc_Zi.Zj *Association between variable Z_i and Z_j.*

Description

Finding association between variables of i^{th} position and j^{th} position. In any position wise aligned sequence dataset, occurences of R=(A,G) and Y=(C, T) at each position can be explained by a standard normal variate Z based on certain threshold value. So, an association between any two position in the datast can be obtained which will be the association beween the two standard normal variate at this two positions. However, the two nomal variates reprsenting the occurences of R and Y are independent of each other at a given position.

Usage

```
assoc_Zi.Zj(x)
```

Arguments

x	A dataframe of position wise aligned sequence dataset having A, T, G and C only.

Details

The user has to supply the sequence dataset in tab delimited format and not in FASTA format. Each sequence (row) should contain only standard nucleotides (A, T, G and C). Each sequence should be same length.

Value

A numeric matrix of order *L* by *L* for the dataset of *L* nucleotides long sequences.

Author(s)

Prabina Kumar Meher & A. R. Rao

Examples

```
data(don_dat)
kk <- don_dat[1:300,]
zizj <- assoc_Zi.Zj(x=kk)
zizj
```

assoc_Zi.ZjR *Association between variable Z_i and Z_jR.*

Description

Finding association between variable Z at i^{th} position and Z_R at j^{th} position. Here, the standard normal variable Z represents the occurence of R=(A,G) and Y=(C, T) at each position in the position wise aligned dataset, whereas the the standard normal variable Z_R reprsents the occurences of nucleotides A and G at any position based on some threshold value.

Usage

```
assoc_Zi.ZjR(x, rZiZj)
```

Arguments

x	A dataframe of position wise aligned sequence dataset having A, T, G and C only.
rZiZj	An object generated by using the function assoc_Zi.Zj.

Details

The user has to supply the input dataset as well as the output generated from the function assoc_Zi.Zj.

Value

A numeric matrix of order *L* by *L* for the dataset of *L* nucleotides long sequences.

Note

It may happen that the convergence will not reach after a ceratin number of iterations and will not produce any output. In such situation, the user is advised to exclude or include some positions , or otherwise include or exclude certain sequences. The user should exploit both options till convergence is reached.

Author(s)

Prabina Kumar Meher & A. R. Rao

Examples

```
data(don_dat)
kk <- don_dat[1:300,]
zizj <- assoc_Zi.Zj(x=kk)
zizjr <- assoc_Zi.ZjR(x=kk, rZiZj=zizj)
zizjr
```

assoc_Zi.ZjY *Association between variable Z_i and Z_jY.*

Description

Finding association between variable Z at i^{th} position and Z_Y at j^{th} position. Here, the standard normal variable Z represents the occurence of R=(A,G) and Y=(C, T) at each position in the position wise aligned dataset, whereas the the standard normal variable Z_R reprsents the occurences of nucleotides A and G at any position based on some threshold values.

Usage

```
assoc_Zi.ZjY(x, rZiZj)
```

Arguments

x	A dataframe of position wise aligned sequence dataset having A, T, G and C only.
rZiZj	An object generated by using the function assoc_Zi.Zj.

Details

The user has to supply the input dataset as well as the output generated from the function assoc_Zi.Zj.

Value

A numeric matrix of order *L* by *L* for the dataset of *L* nucleotides long sequences.

Note

It may happen that the convergence will not reach after a ceratin number of iterations and will not produce any output. In such situation, the user is advised to exclude or include some positions , or otherwise include or exclude certain sequences. The user should exploit both options till convergence is reached.

Author(s)

Prabina Kumar Meher & A. R. Rao

Examples

```
data(don_dat)
kk <- don_dat[1:300,]
zizj <- assoc_Zi.Zj(x=kk)
zizjy <- assoc_Zi.ZjY(x=kk, rZiZj=zizj)
zizjy
```

assoc_ZiR.ZjR *Association between variable Z_iR and Z_jR.*

Description

Finding association between variable Z_R at i^{th} position and Z_R at j^{th} position. Here, the standard normal variable Z_R reprsents the occurences of nucleotides A and G at any position based on some threshold value.

Usage

```
assoc_ZiR.ZjR(x, rZiZj, rZiZjR)
```

Arguments

x	A dataframe of position wise aligned sequence dataset having A, T, G and C only.
rZiZj	An object generated by using the function assoc_Zi.Zj.
rZiZjR	An object generated by using the function assoc_Zi.ZjR.

Details

The user has to supply the input dataset as well as the outputs generated from the functions assoc_Zi.Zj and assoc_Zi.ZjR.

Value

A numeric matrix of order *L* by *L* for the dataset of *L* nucleotides long sequences.

Note

It may happen that the convergence will not reach after a ceratin number of iterations and will not produce any output. In such situation, the user is advised to exclude or include some positions , or otherwise include or exclude certain sequences. The user should exploit both options till convergence is reached.

Author(s)

Prabina Kumar Meher & A. R. Rao

Examples

```
data(don_dat)
kk <- don_dat[1:300,]
zizj <- assoc_Zi.Zj(x=kk)
zizjr <- assoc_Zi.ZjR(x=kk, rZiZj=zizj)
zirzjr <- assoc_ZiR.ZjR(x=kk, rZiZj=zizj, rZiZjR=zizjr)
```

zirzjr

---

assoc_ZiR.ZjY *Association between variable Z_iR and Z_jY.*

Description

Finding association between variable Z_R at i^{th} position and Z_Y at j^{th} position. Here, the standard normal variable Z_Y represents the occurences C and T at each position in the position wise aligned dataset, and the standard normal variable Z_R reprsents the occurences of nucleotides A and G at any position based on some threshold values.

Usage

```
assoc_ZiR.ZjY(x, rZiZj, rZiZjR, rZiZjY)
```

Arguments

x	A dataframe of position wise aligned sequence dataset having A, T, G and C only.
rZiZj	An object generated by using the function assoc_Zi.Zj.
rZiZjR	An object generated by using the function assoc_Zi.ZjR.
rZiZjY	An object generated by using the function assoc_Zi.ZjY.

Details

The user has to supply the input dataset as well as the outputs generated from the functions assoc_Zi.Zj, assoc_Zi.ZjR and assoc_Zi.ZjY.

Value

A numeric matrix of order *L* by *L* for the dataset of *L* nucleotides long sequences.

Note

It may happen that the convergence will not reach after a ceratin number of iterations and will not produce any output. In such situation, the user is advised to exclude or include some positions , or otherwise include or exclude certain sequences. The user should exploit both options till convergence is reached.

Author(s)

Prabina Kumar Meher & A. R. Rao

Examples

```
data(don_dat)
kk <- don_dat[1:300,]
zizj <- assoc_Zi.Zj(x=kk)
zizjr <- assoc_Zi.ZjR(x=kk, rZiZj=zizj)
zizjy <- assoc_Zi.ZjY(x=kk, rZiZj=zizj)
zirzjy <- assoc_ZiR.ZjY(x=kk, rZiZj=zizj, rZiZjR=zizjr, rZiZjY=zizjy)
zirzjy
```

`assoc_ZiY.ZjY` *Association between variable Z_iY and Z_jY.*

Description

Finding association between variable Z_Y at i^{th} position and Z_Y at j^{th} position. Here, the standard normal variable Z_Y reprsents the occurences of nucleotides C and T at any position based on some threshold values.

Usage

```
assoc_ZiY.ZjY(x, rZiZj, rZiZjY)
```

Arguments

`x`	A dataframe of position wise aligned sequence dataset having A, T, G and C only.
`rZiZj`	An object generated by using the function `assoc_Zi.Zj`.
`rZiZjY`	An object generated by using the function `assoc_Zi.ZjY`.

Details

The user has to supply the input dataset as well as the outputs generated from the functions `assoc_Zi.Zj` and `assoc_Zi.ZjY`.

Value

A numeric matrix of order *L* by *L* for the dataset of *L* nucleotides long sequences.

Note

It may happen that the convergence will not reach after a ceratin number of iterations and will not produce any output. In such situation, the user is advised to exclude or include some positions , or otherwise include or exclude certain sequences. The user should exploit both options till convergence is reached.

Author(s)

Prabina Kumar Meher & A. R. Rao

Examples

```
data(don_dat)
kk <- don_dat[1:300,]
zizj <- assoc_Zi.Zj(x=kk)
zizjy <- assoc_Zi.ZjY(x=kk, rZiZj=zizj)
ziyzjy <- assoc_ZiY.ZjY(x=kk, rZiZj=zizj, rZiZjY=zizjy)
ziyzjy
```

---

don_dat *A sample dataset of human donor splice sites.*

Description

This dataset comprises 1000 donor splice site sequences, where each sequence is of length 20 with 10 at the exon end and 10 at the intron start excluding the conserved di-nucleotide GT at the begining of intron. This dataset was randomly taken from true donor splice sites of HS3D dataset.

Usage

```
data(don_dat)
```

References

Pollastro P, Rampone S: HS3D: Homosapiens Splice Site Data Set. *Nucleic Acids Res*. 2003, Molecular Biology Database Collection entry number 36; Annual Database Issue.

Examples

```
data(don_dat)
```

Index