Package ‘DMReSearch’

June 15, 2015

Type Package

Title Detection of Differentially Methylated Regions Based on Three Dimensional Rank Clustering

Version 1.0

Date 2015-06-09

Author Ye Tian and Ruimin Sun

Maintainer Ye Tian <tian0049@e.ntu.edu.sg>

Description

License GPL- (>=3)

LazyData true

R topics documented:

`ClusterMethylC` *Pre-cluster the CpGs by using 3D rank method.*

Description

Pre-cluster the CpGs by using 3D rank method.

Usage

```
ClusterMethylC(methylDataSet, dc = 300, minSize = 3, thCenter = 0.02)
```

Arguments

`methylDataSet`	The well prepared data from MethylDataSet.
`dc`	The maximal distance between two adjacent CpGs in one cluster.
`minSize`	The minimal number of CpGs in one cluster.
`thCenter`	The ratio to choose centers in the 3D rank for each iteration. In each iteration we choose top `thCenter` CpGs as centers.

Value

`methylDataSet`	The updated "methylDataSet" class data.
`clusterRange`	The information of pre-clustered ragions. It includes the start and end positions of all clusters and the number of CpGs in each cluster.

Examples

```
# data("data.toy")
# cluster.toy = ClusterMethylC(data.toy)
```

`data.toy` *The toy data extracted from the Hasen (2011) WGBS data set*

Description

This data set cansists of 1000 CpGs and 6 sample in two biological groups.

Usage

```
data.toy
```

Format

This data set is "MethylDataSet" class data which is well prepared for the following analysis.

References

Hansen, K. D. and Timp, W., Bravo, H. C., Sabunciyan, S., Langmead, B., McDonald, O. G., Wen, B., Wu, H., Liu, Y., Diep, D., Briem, E., Zhang, K., Irizarry, R. A. and Feinberg, A, P. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nature Genetics*, 43, 768-775.

`MethylDataSet` *Filter and generate the "MethylDataSet" class data.*

Description

This function filter CpGs based on their total read counts and prepare the "MethylDataSet" class data for the following analysis. This function can only analyze one chromosome at one time. In order to filter out the CpGs with too low coverage, we only keep the CpGs whose total count is larger than or equal to the `minCount` in at least `filter` times the number of samples in each biological group.

Usage

```
MethylDataSet(sampleDesign, cpgPos, totalCount, methylCount, filter = 2/3,
  minCount = 2)
```

Arguments

sampleDesign	The information of samples. This argument is the bological group labels for samples.
cpgPos	The information of CpGs. It is the list which contain the chromosome label, the strand direction and the position in the strand.
totalCount	The $P \times N$ total count matrix with N samples and P CpGs.
methylCount	The $P \times N$ methylated count matrix with N samples and P CpGs.
filter	The ratio for filter.
minCount	The minimal total count threshold for filter.

Value

y

The "MethylDataSet" class data set.

Examples

```
# sampleDesign = c(rep("cancer", 3), rep("control", 3))
 # names(sampleDesign) = c("C1", "C2", "C3", "N1", "N2", "N3")
 # position = 1:1000
 # cpgPos = list(chr = rep("chr22", 1000), strand = rep("*", 1000), position = position)
 # totalCount = matrix(rpos(6000, lambda = 10), ncol = 6)
 # methylCount = rbinom(6000, size = totalCount, prob = 0.5)
 # toy.data = MethylDataSet(sampleDesign, cpgPos, totalCount, methylCount)
```

SmoothCluster *The modified local kernel smoothing method.*

Description

Smooth the pre-clustered "MethylDataSet" data by using modified local triangular kernel method.

Usage

```
SmoothCluster(object, clusterRange, size = 100)
```

Arguments

object	The "MethlDataSet" data from ClusterMethylC function.
clusterRange	The information of pre-clustered ranges from ClusterMethylC function.
size	The window size for smoothing

Value

smoothed	The matrix that contains the smoothed methylation levels.

Examples

```
# data("data.toy")
# cluster.toy = ClusterMethylC(data.toy)
# smooth.toy = SmoothCluster(cluster.toy$methylDataSet, cluster.toy$clusterRange)
```

TestDMR *Identify and trim the DMRs.*

Description

Based on the wald test this function can detect DMCs and accodingly find the DMRs.

Usage

```
TestDMR(object, smoothed, clusterRange, sigLevel = 0.1, minSize = 3,
  dc = 300, thres = 0.8)
```

Arguments

`object`	The "MethlDataSet" data from `ClusterMethylC` function.
`smoothed`	The smoothed methylation levels from `SmoothCluster`.
`clusterRange`	The information of pre-clustered ragions from `ClusterMethylC` function.
`sigLevel`	The statistical significance level for wald test.
`minSize`	The minimal number of CpGs in one DMR.
`dc`	The maximal distance between two adjacent CpGs in one DMR.
`thres`	The minimal ratio of DMC in the DMR.

Value

`methylDataSet`	The updated "methylDataSet" class data.
`pvalue`	The p-values and the methylation directions of CpGs. For the independent CpGs which cannot be included in any pre-clusters, the corresponding p-values and directions will be NA.
`dmrInfo`	The `data.frame` to keep the DMRs' information. For each DMR, It contains the cluster ID, start position, end position, number of CpGs and mean pvalues.

Examples

```
# data("data.toy")
# cluster.toy = ClusterMethylC(data.toy)
# smooth.toy = SmoothCluster(cluster.toy$methylDataSet, cluster.toy$clusterRange)
# DMR.toy = TestDMR(cluster.toy$methylDataSet, smooth.toy, cluster.toy$clusterRange)
```

Index