Package 'ncdfFlow'

April 5, 2014

Title ncdfFlow: A package that provides ncdf based storage for flow cytometry data.

Version 2.8.22

Author Mike Jiang,Greg Finak,N. Gopalakrishnan

Description Provides netCDF storage based methods and functions for manipulation of flow cytometry data.

Maintainer Mike Jiang <wjiang2@fhcrc.org>

Depends R (>= 2.14.0), flowCore, flowViz

Imports Biobase,flowCore,flowViz,methods,zlibbioc

License Artistic-2.0

SystemRequirements hdf5 (>= 1.8.0)

biocViews FlowCytometry

Collate 'AllGeneric.R' 'AllClasses.R' 'AllFunctions.R' 'ncdfFlowSet-accessors.R' 'ncdfFlowSet-split-methods.R' 'ncdfFlowSet-Subset-methods.R' 'ncdfFlowSet-rbind-methods.R' 'bitVector.R' 'ncdfIO.R' 'ncdfFlowList-methods.R' 'ncdfFlowSet-plot-methods.R' 'Ldpath.R'

R topics documented:

`ncdfFlow-package` *ncdfFlow: A package that provides netCDF storage based flow cytometry data analysis.*

Description

Define important flow cytometry data classes: `ncdfFlowSet`(a subclass of `flowSet`) and `ncdfFlowList`(a list of ncdfFlowSet object) and their accessors.

Provide important compensation,transformation,filter,gating,subsetting,splitting functions for data analysis of large volumns of flow cytometry data that is too big to be held in memory.

Details

Package:	ncdfFlow
Version:	0.99.4
Date:	2011-08-02
Depends:	R (>= 2.8.1), flowCore
License:	Artistic-2.0

Author(s)

Mike Jiang,Greg Finak,N.Gopalakrishnan

Maintainer: Mike Jiang <wjiang2@fhcrc.org>

References

http://www.rglab.org/

`addFrame-methods` *add/replace the data in ncdfFlowSet*

Description

Add one flowFrame to the ncdfFlowSet.(it is defunct, For writing flow data to cdf, use [[<- directly)

Usage

```
addFrame(ncfs,data,sampleName)
```

Arguments

ncfs	a ncdfFlowSet object
data	a flowFrame to be added
sampleName	a character used as sample index of ncdfFlowSet

Details

The dimensions of the flowFrame to be added has to match the target sample data in ncdfFlowSet. It updates the target sample data if it already exists in ncdfFlowSet object.

Author(s)

Mike Jiang,Greg Finak,N.Gopalakrishnan

Maintainer: Mike Jiang <wjiang2@fhcrc.org>

See Also

read.ncdfFlowSet

Examples

```
path<-system.file("extdata","compdata","data",package="flowCore")
files<-list.files(path,full.names=TRUE)[1:3]

##create empty ncdfFlowSet from fcs
nc1 <- read.ncdfFlowSet(files=files,ncdfFile="ncfsTest.nc",flowSetId="fs1",isWriteSlice= FALSE)
fs1<-read.flowSet(files);
#add the actual data slices afterwards
nc1[[1]] <- fs1[[1]]
nc1[[1]]##show the added flowFrame
nc1[[2]]##show empty flowFrame
```

---

clone.ncdfFlowSet *Clone a ncdfFlowSet*

Description

Create a new ncdfFlowSet object from an existing one

Usage

```
clone.ncdfFlowSet(ncfs,ncdfFile=NULL,isEmpty=TRUE,
isNew=TRUE,isSaveMeta=FALSE)
```

Arguments

ncfs	A ncdfFlowSet.
isNew	A logical scalar indicating whether the new cdf file should be created. If FALSE, the original cdf file is associated with the new ncdfFlowSet object.
ncdfFile	A character scalar giving the output file name. By default, It is NULL and the function will generate a random file name, potentially adding the .cdf suffix unless a file extension is already present. It is only valid when isNewNcFile=TRUE
isEmpty	A logical scalar indicating whether the raw data should also be copied.if FALSE, an empty cdf file is created with the same dimensions (sample*events*channels) as the orignial one.
isSaveMeta	A logical scalar indicating whether the meta data other than raw data should be saved in cdf. It should be set as TRUE if the entire ncdfFlowSet is going to be loaded by ncdfFlowSet_open,character-method.

Value

A ncdfFlowSet object

Author(s)

Mike Jiang,Greg Finak,N.Gopalakrishnan

Maintainer: Mike Jiang <wjiang2@fhcrc.org>

See Also

read.ncdfFlowSet

Examples

```
library(ncdfFlow)

path<-system.file("extdata","compdata","data",package="flowCore")
files<-list.files(path,full.names=TRUE)[1:3]

#create ncdfFlowSet from fcs
nc1 <- read.ncdfFlowSet(isSaveMeta=FALSE,files=files,ncdfFile="ncfsTest.nc",flowSetId="fs1",isWriteSlice= TRUE)


##clone the ncdfFlowSet object,by default the actual raw data is not added
nc2<-clone.ncdfFlowSet(nc1,"clone.nc")
nc2[[1]]

#add the actual raw data
fs1  <- read.flowSet(files=files)
nc2[[sampleNames(fs1)[1]]] <- fs1[[1]]
nc2[[1]]


#delete the cdf file associated with ncdfFlowSet before removing it from memory
```

```
ncfsUnlink(nc2)
rm(nc2)

ncfsUnlink(nc1)
rm(nc1)
```

---

ncdfFlowList-class *'ncdfFlowList': a class that stores multiple ncdfFlowSet objects*

Description

It is a list of ncdfFlowSet objects

Objects from the Class

Objects can be created by coercing a list of ncdfFlowSet objects as("ncdfFlowList",nclist = ... #a list of ncdfFlowSet objects)

Methods

show display object summary.

rbind2 combine a list of ncdfFlowSets into one.

lapply The argument 'level' controls whether loop at 'ncdfFlowSet' level or 'sample' level. when level = 2 (default value) FUN is applied to each sample. When level = 1, FUN is applied to each object stored in data slot.

Author(s)

Mike Jiang,Greg Finak,N.Gopalakrishnan

Maintainer: Mike Jiang <wjiang2@fhcrc.org>

See Also

ncdfFlowSet

Examples

```
data(GvHD)
GvHD <- GvHD[pData(GvHD)$Patient %in% 6:7][1:4]
nc1<-ncdfFlowSet(GvHD)

##splitting by a gate
qGate <- quadGate(filterId="qg", "FSC-H"=200, "SSC-H"=400)
fr<-filter(nc1,qGate)
ncList<-split(nc1,fr)
ncList
nc1[1]#each ncdfFlowSet in the list share the same cdf file with the original nc1
nc1[[1]]
```

```
ncList[[2]][[1]]
ncList[[1]][[1]]

##create the new cdf file to detach from its data source
ncList<-split(nc1,fr,isNew=TRUE)
ncList[1]


## split the ncdfFlowSet by a factor
ncList<-split(nc1[1:3], as.factor(c(1,1,2)))
ncList
ncList <- as(ncList, "ncdfFlowList")
summary(ncList[[1]])
sampleNames(ncList)
ncList[[1]]
ncList[, 1:2][[1]]
ncList[["s6a01"]]
ncList["s6a01"]
ncList[c(1,3)][[2]]
ncList[c("s6a01","s6a03"), j= c("FSC-H", "SSC-H")][[1]]
sampleNames(ncList[c(3,1)])
colnames(ncList)
length(ncList)

pData(ncList)
phenoData(ncList)
pData(ncList)$test <- 1
pData(ncList)
xyplot(FSC-H~SSC-H|Visit, ncList)
densityplot(~SSC-H, ncList)
```

`ncdfFlowSet-class` *'ncdfFlowSet': a class for storing flow cytometry raw data in netCDF format*

Description

This class is a subclass of `flowSet`. It stores the raw data in cdf file instead of memory so that the analysis tools provided by flowCore based packages can be used in the study that produces hundreds or thousands FCS files.

Objects from the Class

Objects can be created by using: `read.ncdfFlowSet`,`clone.ncdfFlowSet`

or `ncdfFlowSet( x,#x is a flowSet ncdfFile #the output cdf file name )`

Slots

`file`: A character containing the ncdf file name.

`maxEvents`: An integer containing the maximum number of events of all samples stored in this ncdfFlowSet object

`flowSetId`: A character for the id of ncdfFlowSet object

`indices`: Object of class `"environment"` containing events indices of each sample stored as `"raw"` vector. Each index value is either TURE or FALSE and the entire indices vector is used to subset the raw data. the indices vector of each sample is NA by default when the ncdfFlowSet first created.It is assigned with actual value when ncdfFlowSet object is subsetted by `Subset` or other subsetting methods.

`origSampleVector`: A character vector containing the sample names, which indicates the original order of samples physically stored in cdf format

`origColnames`: A character vector containing the flow channel names, which indicates the original order of columns physically stored in cdf format

`frames`: Object of class `"environment"` which replicates the "frame" slot in `flowSet`, except that `exprs` matrix is empty and the actual data is stored in cdf file.

`phenoData`: see `phenoData`

`colnames`: see `colnames`. Here it serves as the current data view which does not reflect the actual number and order of columns stored in cdf file.

Extends

Class `"flowSet"`, directly.

Methods

Most of the other flowSet methods are not listed here,but they all work as the same due to its inheritance from flowSet.Please see for more `flowSet`details for these methods.

addFrame add or replace the flowFrame in ncdfFlowSet.See `addFrame` for more details

Usage:

```
addFrame(ncfs,data,sampleName)
```

[,[[Subsetting. similar to `[,[[` for flowSet.

[[<- replace data with a `flowFrame`.

Usage:

```
ncfs[[thisSample]] <- value
ncfs[[thisSample,j = c("SSC-A","FSC-A")], check.names = TRUE, only.exprs = FALSE] <- value
```

Note that providing the channel index through `j` can speed up writing process since only the specified channels are written. Also `check.names` flag indicates whether the colnames of flowFrame should be matched to ncdfFlowSet, it can be set as FALSE Thus simply update the first n channels wihtin ncdfFlowSet without matching channel names It is useful in `parseWorkspace` where the flowFrame with pre-fixed colnames needs to be written to ncdfFlowSet where the colnames has not yet ready to be updated in the middle of parsing

`only.exprs` option allows for only updating the exprs data when it is set to TRUE. otherwise, the parameters and decriptions slot within flowFrame are updated as well.

getIndices extract the event indices of one or multiple samples from ncdfFlowSet,return a logical vector.

Usage:

```
getIndices(ncfs,sampleName)
```

initIndices initialize the event indices for the entire ncdfFlowSet with NA

Usage:

```
initIndices(ncfs)
```

updateIndices update the event indices of the target sample in ncdfFlowSet

Usage:

```
updateIndices(ncfs,sampleName,y)
```

ncdfFlowSet_open (Deprecated!)load the ncdfFlowSet object from the cdf file,return a ncdfFlowSet object

Usage:

```
ncdfFlowSet_open(filename)
```

Note that in order to successfully recover the entire ncdfFlowSet object,the phenoData has to be already saved in cdf either by explicitly calling `ncdfFlowSet_sync` or setting `isSaveMeta` as TRUE when ncdfFlowSet is created by `link{read.ncdfFlowSet}` or `link{clone.ncdfFlowSet}`

ncdfFlowSet create ncdfFlowSet from a flowSet object

Usage:

```
ncdfFlowSet(fs1)
```

Note that only flowSet can be coerced to ncdfFlowSet,attempt to apply this method to flowFrame will get an error message.

ncdfFlowSet_sync (Deprecated!)save phenoData to cdf file.

Usage:

```
ncdfFlowSet_sync(filename)
```

ncfsUnlink delete the netCDF file associated with the ncdfFlowSet object

Usage:

```
ncfsUnlink(ncfs)
```

Note that ncdfFlowSet object is unrecoverable after cdf is deleted. So this method is usually called when ncdfFlowSet object is no longer in need.

as.flowSet convert a subset of ncdfFlowSet to flowSet.

Usage:

```
as.flowSet(ncfs,top)
```

Argument `top` specifies the number of samples evenly selected from ncdfFlowSet.

ncfsApply equivalent to `fsApply`., which could cause memory issue due to the. So ncfsApply will return a ncdfFlowSet object.

Usage:

```
ncfsApply(x="ncdfFlowSet",FUN="ANY")
```

Note that the function given by argument "FUN" should return an entire flowFrame object with the same size of the original one(such as compensate,transform...) Otherwise,`fsApply` should be used instead.

rbind2 similar to flowCore:rbind2. It returns a new ncdfFlowSet with a new cdf file that combines two raw datasets. It is recommended to construct a ncdfFlowList and apply rbind2 to it directly when combining more than two ncdfFlowSets. Because using "do.call" on a list ncdfFlowSets will create one cdf file for every pair of ncdfFlowSets,which is not efficient.

split equivalent to flowCore:split. It returns a new ncdfFlowSet object without creating new cdf raw data file but assigning logical indices to subset the original raw data.

Subset equivalent to flowCore:Subset. It returns a new ncdfFlowSet object without creating new cdf raw data file but assigning logical indices to subset the original raw data.

densityplot,xyplot equivalent to flowViz:densityplot and flowViz:xyplot. User need to be careful about applying these plot methods to ncdfFlowSet because it could be slow for large number of flow data.

Author(s)

Mike Jiang,Greg Finak,N.Gopalakrishnan

Maintainer: Mike Jiang <wjiang2@fhcrc.org>

See Also

flowSet,read.ncdfFlowSet,ncdfFlowList

Examples

```
data(GvHD)

nc1 <- ncdfFlowSet(GvHD[1:2])
nc2 <- ncdfFlowSet(GvHD[3:4])
nc3 <- ncdfFlowSet(GvHD[5:6])

##combine two ncdfFlowSets
nc4 <- rbind2(nc1,nc2)

##combine more than two ncdfFlowSets
ncfslist <- as(list(nc1,nc2,nc3),"ncdfFlowList")
nc5 <- rbind2(ncfslist)
```

---

read.ncdfFlowSet *create ncdfFlowSet from FCS files*

Description

read FCS files from the disk and load them into a ncdfFlowSet object

Usage

```
read.ncdfFlowSet(files = NULL,ncdfFile,flowSetId,
isWriteSlice=TRUE,isSaveMeta=FALSE,phenoData,channels=NULL,...)
```

Arguments

files	A character vector giving the source FCS raw file paths.
ncdfFile	A character scalar giving the output file name. Default is NULL and the function will generate a random file in the temporary folder, potentially adding the .cdf suffix unless a file extension is already present. It is sometimes useful to specify this file path to avoid the failure of writing large flow data set to cdf file due to the the shortage of disk space in system temporary folder. It is only valid when isNewNcFile=TRUE
flowSetId	A character scalar giving the unique ncdfFlowSet ID.
isWriteSlice	A logical scalar indicating whether the raw data should also be copied.if FALSE, an empty cdf file is created with the dimensions (sample*events*channels) supplied by raw FCS files.
isSaveMeta	A logical scalar indicating whether the meta data other than raw data should be saved in cdf. It should be set as TRUE if the entire ncdfFlowSet is going to be loaded by ncdfFlowSet_open,character-method.
phenoData	An object of AnnotatedDataFrame providing a way to manually set the phenotyoic data for the whole data set in ncdfFlowSet.
channels	A character vector specifying which channels to extract from FCS files. It can be useful when FCS files do not share exactly the same channel names. Thus this argument is used to select those common channels that are of interests. Default value is NULL and the function will try to scan the FCS headers of all files and determine the common channels.
...	extra arguments to be passed to read.FCS.

Value

A ncdfFlowSet object

Author(s)

Mike Jiang,Greg Finak,N.Gopalakrishnan

Maintainer: Mike Jiang <wjiang2@fhcrc.org>

See Also

clone.ncdfFlowSet

Examples

```
library(ncdfFlow)

path<-system.file("extdata","compdata","data",package="flowCore")
files<-list.files(path,full.names=TRUE)[1:3]

#create ncdfFlowSet from fcs with the actual raw data written in cdf
nc1 <- read.ncdfFlowSet(isSaveMeta=FALSE,files=files,ncdfFile="ncfsTest.nc",flowSetId="fs1",isWriteSlice= TRUE)
nc1
```

```
nc1[[1]]
ncfsUnlink(nc1)
rm(nc1)

#create empty ncdfFlowSet from fcs and add data slices afterwards
nc1  <- read.ncdfFlowSet(files=files,ncdfFile="ncfsTest.nc",flowSetId="fs1",isWriteSlice= FALSE)
fs1<-read.flowSet(files)
nc1[[1]] <- fs1[[1]]
nc1[[1]]
nc1[[2]]
```

Index