Package 'longitudinalDynamics'

September 13, 2021

Type Package

Title Identify Intra-, Inter-donor variations in bulk or single cell longitudinal dataset

Version 0.1.0

Author Suhas Vasaikar [aut, cre],
Aarthi Talla [aut, ctb],
Xiaojun Li [aut, ctb]

Maintainer Suhas Vasaikar <suhas.vasaikar@alleninstitute.org>

Description An implementation of the longitudinal data analysis platform in R. It allows to identify intra-donor, inter-donor variations over longitudinal time points. The analysis can be done on bulk expression dataset without known celltype information or single cell with celltype information or known groups. It allows to identify stable and variable features in given donor and each celltype (or user defined group). The outlier in intra-donor can be performed to identify perturbed or change in sample curresponding to donor/participant. The stable and variable gene signature helps to identify gene signatures in dataset and perturbation specific featureset.

Depends R (>= 3.5.0), methods, grid, graphics, stats, grDevices, knitr

Imports ggplot2 (>= 3.3),
reshape2 (>= 1.4),
tidyverse (>= 1.3),
Seurat (>= 3.9),
ggrepel (>= 0.9),
ComplexHeatmap (>= 2.4),
circlize (>= 0.4),
gridExtra (>= 2.3),
cowplot (>= 1.1),
factoextra (>= 1.0),
SAGx (>= 1.62),
pheatmap (>= 1.0),
Rtsne (>= 0.15),
pbapply (>= 1.4),
lme4 (>= 1.1),
ggforce (>= 0.3)

Suggests ArchR (>= 1.0),
MAST (>= 1.14),
rmarkdown

biocViews Data analysis, Longitudinal data, Single cell, scRNA, scATAC, Software, Visualization

License MIT + file LICENSE

Encoding UTF-8

LazyData true

Roxygen list(markdown = TRUE)

RoxygenNote 7.1.1

VignetteBuilder knitr

R topics documented:

avgExpCalc . . . 2
cvCalcBulk . . . 3
cvCalcSC . . . 3
cvSampleprofile . . . 4
dimUMAPPlot . . . 5
genecircosPlot . . . 6
genecircosPlot_combined . . . 7
genePlot . . . 7
iqrBulk . . . 8
iqrSC . . . 9
lmeVariance . . . 9
longitudinalDynamics . . . 10
longitudinalDynamics_bulk_analysis . . . 12
longitudinalDynamics_scatac_analysis . . . 13
longitudinalDynamics_scrna_analysis . . . 15
scatac_archr_genescore . . . 16
StableFeatures . . . 17
VarFeatures . . . 18

`avgExpCalc` *A avgExpCalc Function*

Description

This function allows you to calculate average gene expression on long-normalized data by group defined by user

Usage

```
avgExpCalc(dataObj, group.by)
```

Arguments

`dataObj`	scRNA object with log-normalized data
`group.by`	Calculate average expression by given group

Examples

```
##Input Expression data
#avgExpCalc(dataObj, group.by)
```

`cvCalcBulk` *A cvCalcBulk Function*

Description

This function allows to calculate Intra-donor variations over longitudinal timepoints. The coefficient of variation (CV) is calculated in Bulk data without group information. CV calculated across samples. It requires longitudinal data matrix/data frame and annotation file.

Usage

```
cvCalcBulk(
  mat = NULL,
  ann = NULL,
  meanThreshold = NULL,
  cvThreshold = NULL,
  housekeeping_genes = NULL,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`mat`	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
`ann`	Annotation table. the must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
`meanThreshold`	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
`cvThreshold`	Coefficient of variation threshold to select variable and stable genes Default is 5 for bulk olink data
`housekeeping_genes`	Optional list of housekeeping genes to focus on. Default is NULL
`fileName`	User-defined file name, Default outputFile
`filePATH`	User-defined output directory PATH Default, current directory

`cvCalcSC` *A cvCalcSC Function*

Description

This function allows to calculate Intra-donor variations over longitudinal timepoints. The coefficient of variation is calculated in single cell data. It requires longitudinal data matrix/data frame and annotation file.

Usage

```
cvCalcSC(
  mat = NULL,
  ann = NULL,
  meanThreshold = NULL,
  cvThreshold = NULL,
  housekeeping_genes = NULL,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

mat	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
ann	Annotation table. the must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame; group, name of the group; group_donor, combined string using group:Sample
meanThreshold	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
cvThreshold	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)
housekeeping_genes	Optional list of housekeeping genes to focus on. Default is NULL
fileName	User-defined file name, Default outputFile
filePATH	User-defined output directory PATH Default, current directory

---

cvSampleprofile *A cvSampleprofile Function*

Description

This function allows to calculate Intra-donor variations over longitudinal timepoints. The coefficient of variation is calculated in single cell data. It requires longitudinal data matrix/data frame and annotation file.

Usage

```
cvSampleprofile(
  mat,
  ann,
  meanThreshold = NULL,
  cvThreshold = NULL,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

mat	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
ann	Annotation table. the must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame; group, name of the group; group_donor, combined string using group:Sample
meanThreshold	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
cvThreshold	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)
fileName	User-defined file name, Default outputFile
filePATH	User-defined output directory PATH Default, current directory

---

dimUMAPPlot *A dimUMAPPlot Function*

Description

This function allows you to perform UMAP visualization of gene of interest list.

Usage

```
dimUMAPPlot(
  ann = NULL,
  rnaObj = NULL,
  countMat = NULL,
  nPC = 30,
  gene_oi = NULL,
  groupName = NULL,
  plotname = NULL,
  filePATH = NULL,
  fileName = NULL
)
```

Arguments

ann	Optional. Annotation table used in case of count/genescore matrix. It must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame; group, name of the group; group_donor, combined string using group:Sample
rnaObj	The seurat scRNA object in case of single cell RNA data (optional).
countMat	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column), in case count matrix for expression data (optional).
nPC	Number of PCAs to be used for UMAP, Default is 30
gene_oi	Genes of interest to explore, required
groupName	User-defined group name column from annotation table or seurat annotation column, required

plotname	User-defined output file name, required
filePATH	User-defined output directory PATH Default, current directory
fileName	User-defined file name, Default outputFile

Examples

```
##Count/genescore matrix data
#dimUMAPPlot(ann=annotation, countMat=countData, nPC=15, gene_oi=var_gene,
#groupName="celltype", plotname="variable", filePATH=filePATH, fileName="ATAC")

##Single cell RNA data
#dimUMAPPlot(rnaObj=SeuratObj, nPC=15, gene_oi=var_gene, groupName="celltype",
#plotname="variable", filePATH=filePATH, fileName="scRNA")
```

---

genecircosPlot *A genecircosPlot Function*

Description

This function allows you to Circos Plot for gene list of interest by group

Usage

```
genecircosPlot(
  data,
  geneList,
  groupColumn = NULL,
  group_oi = NULL,
  colorThreshold = NULL
)
```

Arguments

data	Expression matrix or data frame. Rows represents gene/proteins column represents group:donor (group and donor separated by :)
geneList	Genes of interest to explore
groupColumn	Default 1, use 2 when columns are donor:group format
group_oi	Optional, User-defined groups to consider
colorThreshold	User-defined color threshold in colorspace

Examples

```
##Circos Plot for genes expression in a group
#geneList <- c("IL32","CCL5","TCF7","IL7R","LEF1")
#res <- genecircosPlot(data=cv_res, geneList=geneList)
```

genecircosPlot_combined *A genecircosPlot_combined Function*

Description

This function allows you to Circos Plot for gene list of interest by group

Usage

```
genecircosPlot_combined(
  data,
  geneList,
  groupColumn = NULL,
  group_oi = NULL,
  colorThreshold = NULL
)
```

Arguments

data	Expression matrix or data frame. Rows represents gene/proteins column represents group:donor (group and donor separated by :)
geneList	Genes of interest to explore
groupColumn	Default 1, use 2 when columns are donor:group format
group_oi	Optional, User-defined groups to consider
colorThreshold	User-defined color threshold in colorspace

Examples

```
##Circos Plot for genes expression in a group
#geneList <- c("IL32","CCL5","TCF7","IL7R","LEF1")
#res <- genecircosPlot(data=cv_res, geneList=geneList)
```

genePlot *A genePlot Function*

Description

This function allows you to perform UMAP visualization of gene of interest list.

Usage

```
genePlot(ann = NULL, data = NULL, geneName = NULL, groupName = NULL)
```

Arguments

ann	Optional. Annotation table used in case of count/genescore matrix. It must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame; group, name of the group; group_donor, combined string using group:Sample
data	Average Expression matrix or data frame. Rows represents gene/proteins column represents participant samples with group (optional).
geneName	User-defined gene name
groupName	User-defined group name column from annotation table

Examples

```
#plot <- genePlot(ann=annotation, data=ExpressionData, geneName="FOLR3", groupName="Time")
```

---

iqrBulk *A iqrBulk Function*

Description

This function allows you to perform outlier analysis using defining IQR in bulk dataset. Outlier genes defined as exp > mean +/- 2SD.

Usage

```
iqrBulk(ann, mat, SD_threshold = NULL, fileName = NULL, filePATH = NULL)
```

Arguments

ann	Annotation table. the must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
mat	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
SD_threshold	Standard devaition limit to find outlliers (Eg. SD_threshold= 2, equals to Mean+/- 2SD)
fileName	User-defined file name, Default outputFile
filePATH	User-defined output directory PATH Default, current directory

Examples

```
#filePATH <- getwd()
#IQR_res <- iqrBulk(ann=metadata, mat=datamatrix, fileName="olink", filePATH=filePATH)
```

`iqrSC` *A iqrSC Function*

Description

This function allows you to perform outlier analysis on single cell based average expression matrix by defining IQR. Outlier genes defined as exp > mean +/- 2SD.

Usage

```
iqrSC(
  ann,
  mat,
  groupName = NULL,
  SD_threshold = NULL,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`ann`	Annotation table. the must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
`mat`	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
`groupName`	Group label used to calculate average gene expression and check for outliers
`SD_threshold`	Standard devaition limit to find outlliers (Eg. SD_threshold= 2, equals to Mean+/-2SD)
`fileName`	User-defined file name, Default outputFile
`filePATH`	User-defined output directory PATH Default, current directory

Examples

```
#filePATH <- getwd()
#IQR_res <- iqrSC(ann=metadata, mat=datamatrix, fileName="olink", filePATH=filePATH)
```

`lmeVariance` *A lmeVariance Function*

Description

This function allows you to calculate inter-donor variation between participants over longitudinal time points. It uses linear mixed model to calculate variance contribution from each given feature list

Usage

```
lmeVariance(
  ann,
  mat,
  features,
  meanThreshold = NULL,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

ann	Annotation table. the must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
mat	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
features	Variance analysis carried out for the features provided such as c("PTID", "Time", "Sex")
meanThreshold	Average expression threshold to filter lowly expressed genes/features Default is 0
fileName	User-defined file name, Default outputFile
filePATH	User-defined output directory PATH Default, current directory

Examples

```
##Input Expression data
#filePATH <- getwd()
#lmem_res <- lmeVariance(ann=metadata, mat=datamatrix,
#features=c("PTID", "Time", "Sex"),
#meanThreshold=0.1, fileName="RNA", filePATH=filePATH)
```

---

longitudinalDynamics *A longitudinalDynamics Function*

Description

This function allows you to perform analysis of longitudinal dataset. It requires longitudinal data matrix/data frame and annotation file.

Usage

```
longitudinalDynamics(
  metadata = NULL,
  datamatrix = NULL,
  scrnaObj = NULL,
  atacObj = NULL,
  datatype = NULL,
  omics = NULL,
  fileName,
  features = NULL,
```

```
  meanThreshold = 1,
  cvThreshold = 5,
  coding_genes = NULL,
  avgGroup = NULL,
  housekeeping_genes = NULL,
  group_oi = NULL,
  nPC = 30,
  donorThreshold = NULL,
  groupThreshold = NULL,
  topFeatures = 25,
  method = NULL,
  outputDirectory = NULL
)
```

Arguments

metadata	Annotation table must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
datamatrix	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
scrnaObj	Single cell RNA Seurat object, if datatype is single cell RNA
atacObj	Single cell ATAC genematrix or data frame, if datatype is single cell ATAC
datatype	Data input can be bulk or singlecell
omics	User defined name like RNA, ATAC, Proteomics, FLOW
fileName	User defined filename
features	Variance analysis carried out for the features provided such as c("PTID", "Time", "Sex")
meanThreshold	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
cvThreshold	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)
coding_genes	Selecting protein coding/user-defined gene list only
avgGroup	Group label to be used to calculate average gene expression by group label
housekeeping_genes	Optional list of housekeeping genes to focus on Default is NULL
group_oi	Group of interest to focus on, Default is NULL
nPC	Number of PCAs to be used for UMAP, Default is 30
donorThreshold	Donor threshold number to be used, Default is number of participants
groupThreshold	Group label threshold number to be used, Default is (number of participants x group labels)/2
topFeatures	Number of features to be selected from each group, Default is 25
method	Type of variations to be analyzed, c("intra-donor", "inter-donor", "iqr"), Default is all
outputDirectory	User-defined output directory Default, output

Examples

```
##Bulk data
#load("data/data_Annotation.Rda")
#load("data/Olink_NPX_log2_Protein.Rda")
#longitudinalDynamics(metadata=ann, datamatrix=data, datatype="bulk", omics="Protein",
#fileName="olink", features=c("PTID", "Time", "Sex"), meanThreshold=1, cvThreshold=5,
#outputDirectory="output")

##Single cell RNA data
#load("data/pbmc1-subset.Rda")
#load("data/data_Annotation.Rda")
#metadata=ann
#long_res <- longitudinalDynamics(metadata=ann, scrnaObj=pbmc,
#datatype="singlecell", omics="rna",
#fileName="scrna", features=c("PTID", "Time"),
#meanThreshold=0.1, cvThreshold=10,
#coding_genes=FALSE, avgGroup="celltype",
#housekeeping_genes=housekeeping_genes,
#nPC=15,
#donorThreshold=4, groupThreshold=40, topFeatures=25,
#method=c("intra-donor","inter-donor","iqr"),
#outputDirectory="output")
```

`longitudinalDynamics_bulk_analysis`
A longitudinalDynamics_bulk_analysis Function

Description

This function allows you to perform analysis of longitudinal data from bulk dataset. It requires longitudinal bulk data matrix/data frame at sample level and annotation file.

Usage

```
longitudinalDynamics_bulk_analysis(
  metadata = NULL,
  datamatrix = NULL,
  datatype = NULL,
  omics = NULL,
  fileName = NULL,
  filePATH = NULL,
  features = NULL,
  method = NULL,
  meanThreshold = NULL,
  cvThreshold = NULL
)
```

Arguments

`metadata`	Annotation table must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame

datamatrix	Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)
datatype	Data input can be bulk or singlecell
omics	User defined name like RNA, ATAC, Proteomics, FLOW
fileName	User defined filename
filePATH	User-defined output directory path
features	Variance analysis carried out for the features provided such as c("PTID", "Time", "Sex")
method	Type of variations to be analyzed, c("intra-donor", "inter-donor", "iqr"), Default is all
meanThreshold	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
cvThreshold	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)

Examples

```
##Bulk data
#longitudinalDynamics(metadata=ann, datamatrix=data, datatype="bulk", omics="Protein",
#fileName="olink", features=c("PTID", "Time", "Sex"), meanThreshold=1, cvThreshold=5,
#outputDirectory="output")
```

`longitudinalDynamics_scatac_analysis`
A longitudinalDynamics_scatac_analysis Function

Description

This function allows you to perform analysis of longitudinal dataset from single cell atac based genescore matrix.

Usage

```
longitudinalDynamics_scatac_analysis(
  metadata = NULL,
  genescore = NULL,
  datatype = NULL,
  omics = NULL,
  fileName = NULL,
  filePATH = NULL,
  features = NULL,
  meanThreshold = 0.1,
  cvThreshold = 10,
  coding_genes = NULL,
  avgGroup = NULL,
  housekeeping_genes = NULL,
  group_oi = NULL,
  nPC = NULL,
  method = NULL,
```

```
  donorThreshold = NULL,
  groupThreshold = NULL,
  topFeatures = 25
)
```

Arguments

metadata	Annotation table must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
genescore	Genescore Expression matrix or data frame derived from scATAC. Rows represents genes and column represents participant samples (same as annotation table Sample column)
datatype	Data input can be bulk or singlecell
omics	User defined name like RNA, ATAC, Proteomics, FLOW
fileName	User defined filename
filePATH	User-defined output directory path
features	Variance analysis carried out for the features provided such as c("PTID", "Time", "Sex")
meanThreshold	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
cvThreshold	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)
coding_genes	Selecting protein coding/user-defined gene list only
avgGroup	Group label to be used to calculate average gene expression by group label
housekeeping_genes	Optional list of housekeeping genes to focus on Default is NULL
group_oi	Group of interest to focus on, Default is NULL
nPC	Number of PCAs to be used for UMAP, Default is 30
method	Type of variations to be analyzed, c("intra-donor", "inter-donor", "iqr"), Default is all
donorThreshold	Donor threshold number to be used, Default is number of participants
groupThreshold	Group label threshold number to be used, Default is (number of participants x group labels)/2
topFeatures	Number of features to be selected from each group, Default is 25

Examples

```
##Single cell ATAC data
#longitudinalDynamics(metadata=ann, atacObj=atacObj,
#datatype="singlecell", omics="atac",fileName="scatac",
#features=c("PTID", "Time", "group"),
#meanThreshold=0.1, cvThreshold=10,
#housekeeping_genes=NULL, group_oi=NULL,
#nPC=30, donorThreshold=4, groupThreshold=40, topFeatures=25,
#method=c("intra-donor","inter-donor","iqr"),
#outputDirectory="output")
```

`longitudinalDynamics_scrna_analysis` *A longitudinalDynamics_scrna_analysis Function*

Description

This function allows you to perform analysis of longitudinal dataset from single cell RNA.

Usage

```
longitudinalDynamics_scrna_analysis(
  metadata = NULL,
  dataObj = NULL,
  datatype = "singlecell",
  omics = "scRNA",
  fileName = NULL,
  filePATH = NULL,
  features = NULL,
  meanThreshold = 0.1,
  cvThreshold = 10,
  coding_genes = NULL,
  avgGroup = NULL,
  housekeeping_genes = NULL,
  group_oi = NULL,
  nPC = NULL,
  method = NULL,
  donorThreshold = NULL,
  groupThreshold = NULL,
  topFeatures = 25
)
```

Arguments

`metadata`	Annotation table must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
`dataObj`	Single cell Expression data. The scRNA data should be seurat object. To get more information visit https://cran.r-project.org/web/packages/SeuratObject/SeuratObject.pdf
`datatype`	Default singlecell
`omics`	Default scRNA
`fileName`	User defined filename
`filePATH`	User-defined output directory path
`features`	Variance analysis carried out for the features provided such as c("PTID", "Time", "Sex")
`meanThreshold`	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
`cvThreshold`	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)
`coding_genes`	Selecting protein coding/user-defined gene list only

avgGroup	Group label to be used to calculate average gene expression by group label
housekeeping_genes	Optional list of housekeeping genes to focus on Default is NULL
group_oi	Group of interest to focus on, Default is NULL
nPC	Number of PCAs to be used for UMAP, Default is 30
method	Type of variations to be analyzed, c("intra-donor", "inter-donor", "iqr"), Default is all
donorThreshold	Donor threshold number to be used, Default is number of participants
groupThreshold	Group label threshold number to be used, Default is (number of participants x group labels)/2
topFeatures	Number of features to be selected from each group, Default is 25

Examples

```
##Single cell RNA data
#load("data/pbmc1-subset.Rda")
#load("data/data_Annotation.Rda")
#metadata=ann
#long_res <- longitudinalDynamics(metadata=ann, scrnaObj=pbmc,
#datatype="singlecell", omics="rna",
#fileName="scrna", features=c("PTID", "Time"),
#meanThreshold=0.1, cvThreshold=10,
#coding_genes=FALSE, avgGroup="celltype",
#housekeeping_genes=housekeeping_genes,
#nPC=15,
#donorThreshold=4, groupThreshold=40, topFeatures=25,
#method=c("intra-donor","inter-donor","iqr"),
#outputDirectory="output")
```

---

scatac_archr_genescore

A scatac_archr_genescore Function

Description

This function allows you to calculate genescore matrix from scATAC archR object. This function requires archR package installed and scATAC object created.

Usage

```
scatac_archr_genescore(ArchRProj, groupBy)
```

Arguments

ArchRProj	archR scATAC object for input single cell ATAC longitudinal data
groupBy	Group label to be used to calculate average gene expression by group label, Eg. "celltype"

Examples

```
##Input scATAC data
#genescore <- scatac_archr_genescore(ArchRProj=proj, groupBy="celltype")
```

StableFeatures *A StableFeatures Function*

Description

This function allows you to identify stable genes in a participant across longitudinal timepoints in single cell dataset. The coefficient of variation (CV) obtained from 'cvCalcSC' function used to filter genes/features by CV threshold (cvThreshold). User can identify cvThreshold in different datasets using housekeeping genes CV distribution. The minimum expression of gene (meanThreshold) used to remove lowly expressed genes (spike CV).

Usage

```
StableFeatures(
  ann = NULL,
  group_oi = NULL,
  meanThreshold = NULL,
  cvThreshold = NULL,
  donorThreshold = NULL,
  groupThreshold = NULL,
  topFeatures = 25,
  filePATH = NULL,
  fileName = NULL
)
```

Arguments

ann	Annotation table must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
group_oi	Group of interest to focus on, Default is NULL
meanThreshold	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
cvThreshold	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)
donorThreshold	Donor threshold number to be used, Default is number of participants
groupThreshold	Group label threshold number to be used, Default is (number of participants x group labels)/2
topFeatures	Number of features to be selected from each group, Default is 25
filePATH	User-defined output directory path to load the CV result obtained from cvCalcSC function
fileName	User defined filename

Examples

```
##Single cell RNA data
#stablegene <- StableFeatures(ann=metadata, meanThreshold=0.1, cvThreshold=10,
#donorThreshold=donorThreshold, groupThreshold=groupThreshold,
#topFeatures=25, fileName="scRNA", filePATH=filePATH)
```

VarFeatures *A VarFeatures Function*

Description

This function allows you to identify variable genes in a participant across longitudinal timepoints in single cell dataset. The coefficient of variation (CV) obtained from 'cvCalcSC' function used to filter genes/features by CV threshold (cvThreshold). User can identify cvThreshold in different datasets using housekeeping genes CV distribution. The minimum expression of gene (meanThreshold) used to remove lowly expressed genes (spike CV).

Usage

```
VarFeatures(
  ann = NULL,
  group_oi = NULL,
  meanThreshold = NULL,
  cvThreshold = NULL,
  donorThreshold = NULL,
  groupThreshold = NULL,
  topFeatures = 25,
  filePATH = NULL,
  fileName = NULL
)
```

Arguments

ann	Annotation table must consist column Sample, Participant sample name; PTID, participant; Time, longitudinal time frame
group_oi	Group of interest to focus on, Default is NULL
meanThreshold	Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)
cvThreshold	Coefficient of variation threshold to select variable and stable genes Default is 10 for single cell RNA (100*SD/mean)
donorThreshold	Donor threshold number to be used, Default is number of participants
groupThreshold	Group label threshold number to be used, Default is (number of participants x group labels)/2
topFeatures	Number of features to be selected from each group, Default is 25
filePATH	User-defined output directory path to load the CV result obtained from cvCalcSC function
fileName	User defined filename

Examples

```
#Single cell RNA data
#vargenes <- VarFeatures(ann=metadata, meanThreshold=0.1, cvThreshold=10,
#donorThreshold=donorThreshold, groupThreshold=groupThreshold,
#topFeatures=25, fileName="scRNA", filePATH="output/")
```