Package ‘PALMO’

June 1, 2022

Type Package

Title Identify Intra and Inter-Donor Variations in Bulk or Single Cell
Longitudinal Dataset

Version 0.1.1

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

Description
It is a platform for analyzing longitudinal data from bulk as well as single cell datasets. It allows to identify variations in molecular features within and across donors over longitudinal time points. The analysis can be done on bulk expression dataset without known cell type information or single cell with cell type/user-defined groups. It allows to infer stable and variable features in given donor and each cell type (or user defined group). The outlier analysis can be performed to identify technical/biological perturbed samples in donor/participant. Further, differential analysis can be performed to decipher time-wise changes in gene expression in a cell type.

Depends R (>= 4.0), methods, grid, graphics, stats, grDevices

Imports Seurat (>= 3.9), ggrepel (>= 0.9), pbapply (>= 1.4), lme4 (>=
1.1), ggforce (>= 0.3), MAST (>= 1.14), factoextra (>= 1.0),
Rtsne (>= 0.15), knitr(>= 1.30), dplyr, ggplot2, reshape2,
ComplexHeatmap, circlize, cowplot, pheatmap, tidyverse, utils

Suggests ggpubr, rmarkdown

URL https://github.com/aifimmunology/PALMO

BugReports https://github.com/aifimmunology/PALMO/issues

biocViews GeneExpression, SingleCell, DifferentialExpression

License MIT + file LICENSE

Encoding UTF-8

LazyData false

VignetteBuilder knitr

RoxygenNote 7.1.2

NeedsCompilation no

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

Repository CRAN

Date/Publication 2022-06-01 00:10:02 UTC

R topics documented:

`annotateMetadata` *annotateMetadata Function*

Description

This function allows to add user-defined sample, participant, and time column to a PALMO object in standard format.

Usage

```
annotateMetadata(
  data_object,
  sample_column = "Sample",
  donor_column = "PTID",
  time_column = "Time",
  group_column = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. Contains annotation table and expression matrix or data frame. Rows represent gene/proteins column represents participant samples (same as annotation table Sample column)

`sample_column` Name of Sample column in user input annotation data frame. Default 'Sample'

`donor_column` Name of Donor/participant column in user input annotation data frame. Default 'PTID'

`time_column` Name of Time column in user input annotation data frame. Default 'Time'

`group_column` Optional. Calculate average expression by given group like 'celltype' or 'cluster'

Value

PALMO object

Examples

```
## Not run:
annotateMetadata(data_object=palmo_obj, sample_column='Sample',
donor_column='PTID', time_column='Time')

## End(Not run)
```

`avgExpCalc` *avgExpCalc Function*

Description

This function allows you to calculate average gene expression on log-normalized data by group defined by user. This function uses Seurat function AverageExpression (https://satijalab.org/seurat/reference/averageexpression)

Usage

```
avgExpCalc(data_object, assay = "RNA", group_column)
```

Arguments

`data_object` Input *PALMO* S4 object. Contains annotation table and expression matrix or data frame. Rows represent gene/proteins column represents participant samples (same as annotation table Sample column)

`assay` Single cell data Assay type ('RNA', 'SCT'). Default 'RNA'

`group_column` Calculate average expression by given group like 'celltype' or 'cluster'

Value

PALMO object with avg expression

Examples

```
## Not run:
palmo_obj=avgExpCalc(data_object=palmo_obj, assay='RNA',
group_column='celltype')

## End(Not run)
```

`checkReplicates` *checkReplicates Function*

Description

This function allows you to check for any replicates in data. If present then merge expression of samples by median provided mergeReplicates=TRUE

Usage

```
checkReplicates(data_object, mergeReplicates = FALSE)
```

Arguments

`data_object` Input *PALMO* S4 object. Contains annotation table and expression matrix or data frame. Rows represent gene/proteins column represents participant samples (same as annotation table Sample column)

`mergeReplicates`
Merge replicates expression data by Median. Default FALSE

Value

PALMO object with merged replicates

Examples

```
## Not run:
palmo_obj=checkReplicates(data_object=palmo_obj, mergeReplicates=TRUE)

## End(Not run)
```

createPALMOfromsinglecellmatrix *createPALMOfromsinglecellmatrix Function*

Description

This function allows to create Seurat object from counts and metadata as mentioned in https://search.r-project.org/CRAN/refmans/SeuratObject/html/CreateSeuratObject.html. The seurat object then stored in a newly created PALMO object.

Usage

```
createPALMOfromsinglecellmatrix(data, metadata, anndata = NULL)
```

Arguments

`data` Expression matrix or data frame. Rows represents gene/proteins column represents participant samples (same as annotation table Sample column)

`metadata` Metadata associated with singe cell information. For example rownames are unique cell_barcode and columns are information on each cell_barcode like Sample (source of cell_barcode)

`anndata` Annotation dataframe. It consist of information such as *Sample* (sample name), *PTID* (donor/participant), *Time* (longitudinal timepoints)

Value

PALMO object with scRNA

Examples

```
## Not run:
palmo_obj=createPALMOfromsinglecellmatrix(counts, metadata, annotation)

## End(Not run)
```

createPALMOobject *createPALMOobject Function*

Description

This function allows to create PALMO object using Annotation dataframe and Data dataframe. The Data can be bulk data or single cell data. The bulk input data should consists of rows as genes/proteins/features and column as Sample name (same as user-defined Samples in Annotation dataframe). The single cell data should be Seurat object (please check https://search.r-project.org/CRAN/refmans/SeuratObje
In case Seurat object not available then user can use function createPALMOfromsinglecellmatrix to create PALMO object. The Seurat object/metadata should have Sample column corresponding to Annotation dataframe.

Usage

```
createPALMOobject(anndata, data)
```

Arguments

anndata	Annotation dataframe. It consist of information such as *Sample* (sample name), *PTID* (donor/participant), *Time* (longitudinal timepoints)
data	Data can be bulk data or single cell data

Value

PALMO S4 object

Examples

```
## Not run:
palmo_obj=createPALMOobject(anndata, data)

## End(Not run)
```

cvCalcBulk *cvCalcBulk Function*

Description

This function allows to calculate Intra-donor variations in bulk data over longitudinal timepoints. The coefficient of variation (CV=SD/mean) is calculated in Bulk data in same donor/participant across timepoints.

Usage

```
cvCalcBulk(
  data_object,
  meanThreshold = 1,
  cvThreshold = 5,
  median_cvThreshold = NULL,
  donorThreshold = NULL,
  housekeeping_genes = NULL,
  naThreshold = 1,
  plot_log10 = FALSE,
  selectedFeatures = NULL,
  median_cv_max = NULL,
  plotWidth = 5,
  plotHeight = 8,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. Contains annotation table and expression matrix or data frame. Rows represent gene/proteins column represents participant samples (same as annotation table Sample column)

`meanThreshold` Average expression threshold to filter lowly expressed genes Default is 1 (log2 scale)

`cvThreshold` Coefficient of variation threshold to select variable and stable genes. Default is 5 for bulk data. Users can use 10-20 for single cell average expression data.

`median_cvThreshold`
Optional, median of CVs from each donor/participant calculated. Threshold used to differentiate variable and stable features across donors/participants. Default, same as *cvThreshold*.

`donorThreshold` Donor threshold number to be used, Default is number of participants

`housekeeping_genes`
Optional, vector of housekeeping genes. Default is c("ACTB", "GAPDH")

`naThreshold` Optional, For a give feature % of donors/participants showing non-NA CVs (NAs appear due to expression ~0 or absent). Default is 1 means all donors/participants to consider. 0.5 means from 4 donors atleast 2 donors should have non-NA CVs for a given feature.

`plot_log10` Optional, Plot CV vs Mean on log10 scale. Default FALSE

`selectedFeatures`
Optional, focus on selected genes/features.

`median_cv_max` Optional, Remove features with greater than median CV Default is NULL

`plotWidth` Optional, heat plot width 5 in

`plotHeight` Optional, heat plot height 8 in

`fileName` User-defined file name, Default outputFile

`filePATH` User-defined output directory PATH Default, current directory

Value

PALMO object with CV list

Examples

```
## Not run:
palmo_obj=cvCalcBulk(data_object=palmo_obj, meanThreshold=0.1, cvThreshold=5)

## End(Not run)
```

cvCalcBulkProfile *cvCalcBulkProfile Function*

Description

This function allows to calculate Intra-donor variations in bulk data over longitudinal timepoints and visualize in a CV vs Mean plot. Plots stored in output directory.

Usage

```
cvCalcBulkProfile(data_object, cl = 2, fileName = NULL, filePATH = NULL)
```

Arguments

`data_object`	Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.
`cl`	Number of clusters. Use nCores-1 to run parallel. Default 2
`fileName`	User-defined filename, Default outputFile
`filePATH`	User-defined output directory *PATH* Default, current directory

Value

PALMO object with CV profile cv_all

Examples

```
## Not run:
cvCalcBulkProfile(data_object=palmo_obj)

## End(Not run)
```

cvCalcSC *cvCalcSC Function*

Description

This function allows to calculate Intra-donor variations in single cell data over longitudinal timepoints. The coefficient of variation (CV=SD/mean) is calculated in average expression data in same donor/participant and corresponding user-defined group (like celltype, cluster) across longitudinal timepoints.

Usage

```
cvCalcSC(
  data_object,
  meanThreshold = NULL,
  cvThreshold = NULL,
  housekeeping_genes = NULL,
  cl = 2,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. Contains annotation table and expression matrix or data frame. Rows represent gene/proteins column represents participant samples (same as annotation table Sample column)

`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 data. Users can use 10-20 for single cell average expression data.

`housekeeping_genes` Optional, vector of housekeeping genes. Default is c('ACTB', 'GAPDH')

`cl` Number of clusters. Use nCores-1 to run parallel. Default 2

`fileName` User-defined file name, Default outputFile

`filePATH` User-defined output directory *PATH* Default, current directory

Value

PALMO object with CV list

Examples

```
## Not run:
palmo_obj=cvCalcSC(data_object=palmo_obj, meanThreshold=0.1, cvThreshold=5)

## End(Not run)
```

`cvCalcSCProfile` *cvCalcSCProfile Function*

Description

This function allows to calculate Intra-donor variations in single cell data over longitudinal timepoints and visualize in a CV vs Mean plot. Plots stored in output directory.

Usage

```
cvCalcSCProfile(
  data_object,
  meanThreshold = NULL,
  housekeeping_genes = NULL,
  cl = 2,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. Contains annotation table and expression matrix or data frame. Rows represent gene/proteins column represents participant samples (same as annotation table Sample column)

`meanThreshold` Average expression threshold to filter lowly expressed genes Default is 0.1 (log2 scale)

`housekeeping_genes` Optional, vector of housekeeping genes. Default is c('ACTB', 'GAPDH')

`cl` Number of clusters. Use nCores-1 to run parallel. Default 2

`fileName` User-defined file name, Default outputFile

`filePATH` User-defined output directory *PATH* Default, current directory

Value

PALMO object with CV profile list

Examples

```
## Not run:
palmo_obj <- cvCalcSCProfile(data_object=palmo_obj,
housekeeping_genes=c('GAPDH', 'ACTB'), fileName='scrna')

## End(Not run)
```

`cvSCsampleprofile` *cvSCsampleprofile Function*

Description

This function allows to calculate Intra-donor variations in single cell data at sample level over longitudinal timepoints and visualize in a CV vs Mean plot. Plots stored in output directory.

Usage

```
cvSCsampleprofile(
  data_object,
  meanThreshold = NULL,
  cvThreshold = NULL,
  cl = 2,
  plot_log10 = FALSE,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object`	Input *PALMO* S4 object. Contains annotation table and expression matrix or data frame. Rows represent gene/proteins column represents participant samples (same as annotation table Sample column)
`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)
`cl`	Number of clusters. Use nCores-1 to run parallel. Default 2
`plot_log10`	Optional, Plot CV vs Mean on log10 scale. Default FALSE
`fileName`	User-defined file name, Default outputFile
`filePATH`	User-defined output directory *PATH* Default, current directory

Value

PALMO object with CV list

Examples

```
## Not run:
palmo_obj <- cvSCsampleprofile(data_object=palmo_obj,
housekeeping_genes=c('GAPDH', 'ACTB'), fileName='scrna')

## End(Not run)
```

`dimUMAPPlot` *dimUMAPPlot Function*

Description

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

Usage

```
dimUMAPPlot(
  data_object,
  nPC = 30,
  gene_oi,
  group_column,
  plotname = NULL,
  repel = FALSE,
  filePATH = NULL,
  fileName = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. Contains annotation table and single cell data stored as Seurat scRNA object.

`nPC` Number of PCAs to be used for UMAP, Default is 30

`gene_oi` Genes of interest to explore, required

`group_column` User-defined group name column from annotation table or Seurat annotation column. Example, group_column='celltype' (required)

`plotname` User-defined output file name (required)

`repel` UMAP plot with labels repel=TRUE. Default FALSE

`filePATH` User-defined output directory *PATH* Default, current directory

`fileName` User-defined file name, Default outputFile

Value

UMAP plot

Examples

```
## Not run:
dimUMAPPlot(data_object=pamo_obj, nPC=15, gene_oi=stable_gene,
group_column='celltype', plotname='stable')

## End(Not run)
```

`genecircosPlot` *genecircosPlot Function*

Description

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

Usage

```
genecircosPlot(
  data = NULL,
  data_object = NULL,
  geneList,
  group_position = 1,
  group_oi = NULL,
  titleName = "",
  colorThreshold = 10,
  colorMax = NULL,
  colorscale = FALSE
)
```

Arguments

`data` Expression matrix or data frame. Rows represents gene/proteins column represents group:donor (or donor:group)

`data_object` Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data. Rows represents gene/proteins column represents group:donor (or donor:group)

`geneList` User-defined genes of interest

`group_position` Default 1, use 2 when columns are donor:group format

`group_oi` Optional, User-defined groups to consider and order plot

`titleName` Title of the plot

`colorThreshold` User-defined color threshold (same as cvThreshold, like 5)

`colorMax` Maximum CV value in heatplot ("max", numeric or NULL)

`colorscale` Show color scale, TRUE or FALSE (default).

Value

Circos plots and dataframe

Examples

```
## Not run:
genecircosPlot(data_object=palmo_obj, geneList=c('IL32','CCL5','TCF7'))

## End(Not run)
```

gene_featureplot *gene_featureplot Function*

Description

This function allows to output the user-defined input features expression in graphical format. Users can select x-axis as donor/participant (x_group_by='PTID') and expression on y-axis organized by variable time (var_oi='Time'). Add group facet feature like facet_by='celltype'.

Usage

```
gene_featureplot(
  data_object = NULL,
  data = NULL,
  anndata = NULL,
  featureList,
  x_group_by = "PTID",
  var_oi = "Time",
  xlab = "group_by",
  ylab = "Value/Expression",
  ncol = NULL,
  facet_by = NULL,
  compare_means = FALSE,
  x_text_angle = NULL,
  text_font = NULL
)
```

Arguments

`data_object`	Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.
`data`	Optional, Data can be bulk data or single cell data
`anndata`	Optional, Annotation dataframe consist of information such as Sample (sample name), PTID (donor/participant), Time (longitudinal timepoints)
`featureList`	User-defined feature or genelist as a vector
`x_group_by`	x-axis grouping variable like 'PTID'
`var_oi`	x-axis subgrouping variable like 'Time'
`xlab`	x-axis label
`ylab`	y-axis label
`ncol`	Number of columns in the plot grid
`facet_by`	A set of variables or expressions
`compare_means`	Add mean comparison p-value in a plot (for more information refer http://rpkgs.datanovia.com/ggpubr/ref
`x_text_angle`	xaxis text angle on ggplot
`text_font`	font size on ggplot

Value

gene plot

Examples

```
## Not run:
plots <- gene_featureplot(data_object=palmo_obj,
featureList=c('LILRA4', 'CLEC9A'))

## End(Not run)
```

`lmeVariance` *lmeVariance Function*

Description

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

Usage

```
lmeVariance(
  data_object,
  featureSet,
  fixed_effect_var = NULL,
  meanThreshold = NULL,
  selectedFeatures = NULL,
  NA_to_zero = FALSE,
  cl = 2,
  lmer_control = FALSE,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.

`featureSet` Variance analysis carried out for the feature set provided such as c('PTID', 'Time', 'Sex')

`fixed_effect_var` Fixed effect variables. In linear mixed model fixed_effect_var included as fixed effect variables and variance contribution obtained by adding them as random variables

`meanThreshold` Average expression threshold to filter lowly expressed genes/features Default is 0

`selectedFeatures`
User-defined gene/feature list

`NA_to_zero` Convert NAs to zero. Default FALSE

`cl` Number of clusters. Use nCores-1 to run parallel. Default 2

`lmer_control` control structures for mixed model fitting. Default optimizer is "bobyqa". Reduces the run time for large data significantly.

`fileName` User-defined file name, Default outputFile

`filePATH` User-defined output directory *PATH* Default, current directory

Value

PALMO object with variance lmem_res dataframe

Examples

```
## Not run:
palmo_obj=lmeVariance(data_object=palmo_obj,
featureSet=c('PTID','Time','Sex'))

## End(Not run)
```

`longitudinalmfuzz` *longitudinalmfuzz Function*

Description

This function allows you to identify gene/feature trajectory over longitudinal points. The function uses mfuzz package (for more information refer to https://www.bioconductor.org/packages/release/bioc/html/Mfuzz.html)

Usage

```
longitudinalmfuzz(
  data_object,
  group_column = "group",
  timeColumn = "Time",
  timeOrder = NULL,
  donorColumn = "PTID",
  baseline_timepoint = NULL,
  featurelist = NULL,
  group_oi = NULL,
  mfuzz_thres = 0.25,
  mfuzz_min.std = 0,
  max_cluster = NULL,
  delta = 0.5,
  plotsize = 10,
  cl = 2,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object`	Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.
`group_column`	User-defined group name like 'group','celltype'
`timeColumn`	User-defined time column name like 'Time'
`timeOrder`	(Optional) User-defined order of time variable like ('D1','D2','D3')
`donorColumn`	User-defined donor/participant column name like 'PTID'
`baseline_timepoint`	(Optional) If baseline donors known (like 'PTID1')
`featurelist`	(Optional) User-defined genes/features of interest
`group_oi`	User-defined groups to consider for example from celltypes select few
`mfuzz_thres`	`mfuzz:thres` threshold for excluding genes
`mfuzz_min.std`	`mfuzz:min.std` threshold for minimum standard deviation
`max_cluster`	Number of clusters to explore (Default 2^n)
`delta`	`mfuzz:delta` threshold for minimum standard deviation
`plotsize`	Size of plot width and height. Default 10 (in).
`cl`	Number of clusters. Use nCores-1 to run parallel. Default 2
`fileName`	User-defined file name, Default outputFile
`filePATH`	User-defined output directory *PATH* Default, current directory

Value

longitudinal trajectory dataframe

Examples

```
## Not run:
longitudinalmfuzz(data_object=palmo_obj, group_column='group',
timeColumn='Time', donorColumn='PTID')

## End(Not run)
```

`mergePALMOdata` *mergePALMOdata Function*

Description

This function allows to merge expression data from bulk or single cell data with annotation file provided by user. It will remove the annotations with missing information from Sample name, donor/participant and time variable.

Usage

```
mergePALMOdata(data_object, datatype)
```

Arguments

`data_object`	Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.
`datatype`	Input data type 'bulk' or 'singlecell'

Value

PALMO object with merged annotation and data matrix

Examples

```
## Not run:
palmo_obj <- mergePALMOdata(data_object=palmo_obj)

## End(Not run)
```

`multimodalView` *multimodalView Function*

Description

This function allows to visualize the multimodal view genes of interest by celltypes/ groups defined by use

Usage

```
multimodalView(
  modality1,
  modality2,
  group_oi = NULL,
  geneList,
  colorThreshold = 10,
  group_position = NULL,
  plotHeight = 10,
  titleName = "",
  colorMax = NULL,
  colorscale = FALSE,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`modality1`	Variation or Expression matrix/data frame. Rows represents gene/proteins column represents group:donor (group and donor separated by :)
`modality2`	Variation or Expression matrix/data frame. Rows represents gene/proteins column represents group:donor (group and donor separated by :)

`group_oi`	Optional, User-defined groups to consider and order
`geneList`	Genes of interest to explore
`colorThreshold`	User-defined color threshold in color space
`group_position`	Default 1, use 2 when columns are donor:group format
`plotHeight`	User-defined Plot size (in)
`titleName`	Title of the plot
`colorMax`	Maximum CV value in heatplot ("max", numeric or NULL)
`colorscale`	Show color scale, TRUE or FALSE (default).
`fileName`	User defined filename
`filePATH`	User-defined output directory *PATH* to save result

Value

Multimodal plot and data list

Examples

```
## Not run:
multimodalView(modality1=scrna_cv_res, modality2=scatac_cv_res, geneList)

## End(Not run)
```

`naFilter` *naFilter Function*

Description

This function allows users to filter genes/features having expression NAs with user-defined threshold.

Usage

```
naFilter(data_object, na_cutoff = 0.4)
```

Arguments

`data_object`	Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.
`na_cutoff`	Threshold to remove NAs. Range is 0-1 (Default 0.4 means40% NAs are not allowed).

Value

PALMO object with filtered NAs

Examples

```
## Not run:
palmo_obj <- naFilter(data_object=palmo_obj, na_cutoff=0.4)

## End(Not run)
```

outlierDetect *outlierDetect Function*

Description

This function allows users to perform outlier analysis on bulk data by calculating z-score. Outlier genes defined as mean/SD = |Z| > z_cutoff.

Usage

```
outlierDetect(
  data_object,
  z_cutoff = NULL,
  plotWidth = 10,
  plotHeight = 5,
  group_column = NULL,
  cl = 2,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.

`z_cutoff` |Z| cutoff threshold to find potential outliers (Eg. *z_cutoff*=2, equals to `Mean/SD` 2)

`plotWidth` User-defined plot width, Default 10 in

`plotHeight` User-defined plot height, Default 5 in

`group_column` Include group by outlier analysis (celltype, cluster)

`cl` Number of clusters. Use nCores-1 to run parallel. Default 2

`fileName` User-defined file name, Default outputFile

`filePATH` User-defined output directory *PATH* Default, current directory

Value

PALMO object with outlier_res dataframe

Examples

```
## Not run:
palmo_obj <- outlierDetect(data_object=palmo_obj, z_cutoff=2)

## End(Not run)
```

---

outlierDetectP *outlierDetectP Function*

Description

This function allows to identify significant abnormal event identified from outlier analysis.

Usage

```
outlierDetectP(
  outlier_events,
  z_cutoff = 2,
  nGenes,
  group_by = "PTID",
  alternative = "two.sided"
)
```

Arguments

`outlier_events`	Identified outlier events
`z_cutoff`	|Z| cutoff threshold to find potential outliers (Eg. z_cutoff= 2, equals to Mean/SD 2)
`nGenes`	Number of background genes/features
`group_by`	Column name to use for groupwise outlier analysis. Default is PTID (donor or participant id).
`alternative`	alternative hypothesis, must be one of "one.sided" or "two.sided" (default)

Value

PALMO object with outlier event p value dataframe

PALMO object with outlier event p value

Examples

```
## Not run:
outlierDetectP(outlier_events=outlier_res, z_cutoff=2, nGenes=1043)

## End(Not run)
```

palmo-class *palmo class*

Description

This function creates *PALMO* class object. All the raw data and results from PALMO are stored in this object.

Value

PALMO S4 class

Fields

`raw` list, contains user entered annotation and expression dataframe or object

`curated` list, contains curated input data

`result` list, output from *PALMO* stored in result list

`nDonors` numeric, number of donors in the input data

`rownames` character, row names of the expression data

`colnames` character, column names of the expression data

`housekeeping_genes` character, user-defined housekeeping genes listed

`datatype` character, datatype used like bulk or singlecell

`omics` character, omics such as RNA, scRNA, scATAC

`featureSet` character, parameters used for variance analysis

`meanThreshold` numeric, Average expression threshold

`cvThreshold` numeric, CV threshold

`median_cvThreshold` numeric, median of CV threshold (from inter-donor)

`groupName` character, group defined by user like celltype, cluster

`group_oi` character, selected types from user-defined group list

`donorThreshold` numeric, minimum donors to explore

`groupThreshold` numeric, minimum group types to explore

`topFeatures` numeric, number of top features to retrieve

`donor_sep` character, donor and group separator such as ':'

`cor_method` character, correelation method 'pearson', 'spearman'

`clusterBy` character, cluster by a group (celltype or cluster)

`z_cutoff` numeric, z-cutoff value for outlier analysis

`filePATH` character, PATH of outout directory

p_value_for_event *p_value_for_event Function*

Description

This function allows to calculate p value for identified outlier significant abnormal events

Usage

```
p_value_for_event(events, tries, rate)
```

Arguments

events	Identified outlier events
tries	Number of background genes/features
rate	probability distribution

Value

outlier event p value

Examples

```
## Not run:
p_value_for_event(events, tries, rate)

## End(Not run)
```

sample_correlation *sample_correlation Function*

Description

This function allows to perform sample correlation (by group like celltype, or by donor).

Usage

```
sample_correlation(
  data_object,
  cor_method = "spearman",
  group_by = NULL,
  col_break = NULL,
  col_max = 1,
  cluster_rows = FALSE,
  cluster_columns = FALSE,
  column_names_fontsize = 4,
  row_names_fontsize = 4,
  row_title_fontsize = 6,
  column_title_fontsize = 6,
  plotHeight = 20,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.

`cor_method` (Optional) Correlation method 'pearson' or 'spearman'. Default is 'spearman'

`group_by` Cluster correlation heat plot by 'donor' or 'group'

`col_break` Value between 0 and 1

`col_max` Maximum color limit (Default 1)

`cluster_rows` *ComplexHeatmap* cluster rows, Default FALSE

`cluster_columns` *ComplexHeatmap* cluster columns, Default FALSE

`column_names_fontsize` Font size of the column names, Default 4

`row_names_fontsize` Font size of the row names, Default 4

`row_title_fontsize` Font size of the row title, Default 6

`column_title_fontsize` Font size of the column title, Default 6

`plotHeight` Height of the plot in inch, Default 20 in

`fileName` User-defined file name, Default outputFile

`filePATH` User-defined output directory *PATH* Default, current directory

Value

PALMO object with correlation cor_res dataframe

Examples

```
## Not run:
palmo_obj <- sample_correlation(data_object=palmo_obj, group_by="Time")

## End(Not run)
```

`sclongitudinalDEG` *sclongitudinalDEG Function*

Description

This function allows ser to calculate differential expressed genes in the direction of given time points (if timepoints>3 otherwise DEGs between two timepoints). A hurdle model was fit to each participant independently in order to identify participant-specific longitudinal transcriptomic changes. Genes that were expressed in at least 10% of cells per participant were considered for this analysis. The models were fit on the input normalized data, modeling the timepoints as a continuous variable within each cell type and adjusting for the batch only if any timepoints from the same participant were run across multiple batches.

Usage

```
sclongitudinalDEG(
  data_object,
  scassay = "RNA",
  group_column,
  group_oi = NULL,
  mincellsexpressed = 0.1,
  removelnc = "TRUE",
  adjfac = NULL,
  baseline = NULL,
  addCDR = FALSE,
  CDR_column = NULL,
  plotWidth = 10,
  plotHeight = 10,
  fileName = NULL,
  filePATH = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.

`scassay` Single cell assay from scRNA seurat object (Default "RNA")

`group_column` Column of interest such as "celltype" to analyze DEGs in participant over time

`group_oi` Features of interest such as specific celltypes c("CD4_Naive", "CD4_TEM")

`mincellsexpressed`
Average expression threshold to filter lowly expressed genes/features Default is 0.1

`removelnc` Remove lincRNAs, mitochondrial and ribosomal genes from analysis incldes (^RP|^MT-|^LINC|orf) (TRUE/FALSE). Default is TRUE

`adjfac` Factors to be adjusted for such as batch, sex

`baseline`	Donors (PTID) to be considered as baseline. Deafult NULL
`addCDR`	(Optional) Add CDR while performing differential analysis. Default is FALSE
`CDR_column`	(Optional) cellular detection rate column name
`plotWidth`	User-defined plot width, Default 10 in
`plotHeight`	User-defined plot height, Default 10 in
`fileName`	User-defined file name, Default outputFile
`filePATH`	User-defined output directory *PATH* Default, current directory

Examples

```
## Not run:
palmo_obj <- sclongitudinalDEG(ann=metadata, dataObj=pbmc, scassay="RNA",
group_column="celltype")

## End(Not run)
```

`StableFeatures` *StableFeatures Function*

Description

This function allows user to identify stable genes in participants across longitudinal timepoints using single cell expression data. The coefficient of variation (CV) calculated using `cvCalcSC` function. Users can identify cvThreshold in different datasets using housekeeping genes CV distribution.

Usage

```
StableFeatures(
  data_object,
  group_oi = NULL,
  cvThreshold = NULL,
  donorThreshold = NULL,
  housekeeping_genes = NULL,
  groupThreshold = NULL,
  topFeatures = 25,
  filePATH = NULL,
  fileName = NULL
)
```

Arguments

`data_object`	Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.
`group_oi`	Group of interest to focus on. Example among celltypes focus on selected ones. Default is NULL.

`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

`housekeeping_genes`
Optional list of housekeeping genes to focus on. Default is ACTB, GAPDH

`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

Value

PALMO object with stable (stable_genes) features

Examples

```
## Not run:
palmo_obj <- StableFeatures(data_object=palmo_obj, cvThreshold=10)

## End(Not run)
```

VarFeatures *VarFeatures Function*

Description

This function allows user to identify variable genes in participants across longitudinal timepoints using single cell expression data. The coefficient of variation (CV) calculated using `cvCalcSC` function. Users can identify cvThreshold in different datasets using housekeeping genes CV distribution.

Usage

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

Arguments

data_object Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.

group_oi Group of interest to focus on. Example among celltypes focus on selected ones. Default is NULL.

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 *CV* result obtained from cvCalcSC function

fileName User defined filename

Value

PALMO object with variable (var_genes) features

Examples

```
## Not run:
palmo_obj <- VarFeatures(data_object=palmo_obj, cvThreshold=10)

## End(Not run)
```

variancefeaturePlot *variancefeaturePlot Function*

Description

This function allows user to plot variance observed in the data by provided featureList

Usage

```
variancefeaturePlot(
  data_object = NULL,
  vardata = NULL,
  featureSet = "PTID",
  Residual = FALSE,
  top_n = 15,
  cols = NULL,
  ncol = NULL
)
```

Arguments

`data_object` Input *PALMO* S4 object. It contains annotation information and expression data from Bulk or single cell data.

`vardata` Variance result obtained from lmeVariance function

`featureSet` Column of interest to focus on, Default is 'PTID'

`Residual` Add residual in plot, Default FALSE

`top_n` Number of top features to show. Default is 10.

`cols` The colors associated with features. Default is NULL.

`ncol` Plot_grid number of plot columns.

Value

variance plot list

Examples

```
## Not run:
variancefeaturePlot(data_object=palmo_obj, top_n=15)

## End(Not run)
```

Index