Package 'SlimR'

August 19, 2025

Title Marker-Based Package for Single-Cell and Spatial-Transcriptomic Annotation

Version 1.0.7

Description
Annotating single-cell and spatial-transcriptomic (ST) data based on the Marker dataset. It supports the creation of a unified marker list, Markers_list, using sources including: the package's built-in curated species-specific cell type and marker reference databases (e.g., 'Cellmarker2', 'PanglaoDB', 'scIBD', 'TCellSI'), Seurat objects containing cell label information, or user-provided Excel tables mapping cell types to markers. Based on the Markers_list, 'SlimR' can calculate gene expression of different cell types and predict annotation information and calculate corresponding AUC by 'Celltype_Calculate()', and annotate it by 'Celltype_Annotation()', then verify it by 'Celltype_Verification()'. At the same time, it can calculate gene expression corresponding to the cell type to generate the corresponding annotation reference map for manual annotation (e.g., 'Heatmap', 'Features plot', 'Combined plot'). For more details see Kabacoff (2020, ISBN:9787115420572).

License MIT + file LICENSE

Date 2025-08-19

Encoding UTF-8

RoxygenNote 7.3.2

Depends R (>= 3.5)

LazyData true

Imports cowplot, dplyr, ggplot2, patchwork, pheatmap, readxl, scales, Seurat, tidyr, tools, tibble

Suggests crayon

NeedsCompilation no

Author Zhao qing Wang [aut, cre] (ORCID: <https://orcid.org/0000-0001-8348-7245>)

Maintainer Zhao qing Wang <zhaoqingwang@mail.sdu.edu.cn>

Repository CRAN

Date/Publication 2025-08-18 23:30:02 UTC

Contents

`calculate_expression` *Counts average expression of gene set (Use in package)*

Description

Counts average expression of gene set (Use in package)

Usage

```
calculate_expression(
  object,
  features,
  assay = NULL,
  cluster_col = NULL,
  colour_low = "white",
  colour_high = "navy"
)
```

Arguments

object	Enter a Seurat object.
features	Enter one or a set of markers.
assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = NULL".
cluster_col	Enter the meta.data column in the Seurat object to be annotated, such as "seurat_cluster". Default parameters use "cluster_col = NULL".
colour_low	Color for lowest expression level. (default = "white")
colour_high	Color for highest expression level. (default = "black")

Value

Average expression genes and relatied informations in the input "Seurat" object given "cluster_col" and given "features".

See Also

Other Use_in_packages: `calculate_probability()`

---

calculate_probability *Calculate gene set expression and infer probabilities with control datasets (Use in package)*

Description

Calculate gene set expression and infer probabilities with control datasets (Use in package)

Usage

```
calculate_probability(
  object,
  features,
  assay = NULL,
  cluster_col = NULL,
  min_expression = 0.1,
  specificity_weight = 3
)
```

Arguments

object	Enter a Seurat object.
features	Enter one or a set of markers.
assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = NULL".

`cluster_col` Enter the meta.data column in the Seurat object to be annotated, such as "seurat_cluster". Default parameters use "cluster_col = NULL".

`min_expression` The min_expression parameter defines a threshold value to determine whether a cell's expression of a feature is considered "expressed" or not. It is used to filter out low-expression cells that may contribute noise to the analysis. Default parameters use "min_expression = 0.1".

`specificity_weight` The specificity_weight parameter controls how much the expression variability (standard deviation) of a feature within a cluster contributes to its "specificity score." It amplifies or suppresses the impact of variability in the final score calculation.Default parameters use "specificity_weight = 3".

Value

Average expression of genes in the input "Seurat" object given "cluster_col" and given "features".

See Also

Other Use_in_packages: `calculate_expression()`

`Cellmarker2` *Cellmarker2 dataset*

Description

A dataset containing marker genes for different cell types from Cellmarker2

Usage

```
Cellmarker2
```

Format

A data frame with 8 columns:

Details

This dataset is used to filter and create a standardized marker list. The dataset can be filtered based on species, tissue class, tissue type, cancer type, and cell type to generate a list of marker genes for specific cell types.

Source

http://117.50.127.228/CellMarker/

See Also

Other SlimR_Database: `Cellmarker2_raw`, `Cellmarker2_table`, `Markers_list_TCellSI`, `Markers_list_scIBD`, `PanglaoDB`, `PanglaoDB_raw`, `PanglaoDB_table`

Cellmarker2_raw *Cellmarker2 raw dataset*

Description

A dataset containing marker genes for different cell types from Cellmarker2

Usage

```
Cellmarker2_raw
```

Format

A data frame with 20 columns contined in the Cellmarker2 database:

Details

This dataset is used to filter and create a standardized marker list. The dataset can be filtered based on species, tissue class, tissue type, cancer type, and cell type to generate a list of marker genes for specific cell types.

Source

http://117.50.127.228/CellMarker/

See Also

Other SlimR_Database: Cellmarker2, Cellmarker2_table, Markers_list_TCellSI, Markers_list_scIBD, PanglaoDB, PanglaoDB_raw, PanglaoDB_table

Cellmarker2_table *Cellmarker2 table*

Description

A dataset containing marker genes for different cell types from Cellmarker2

Usage

```
Cellmarker2_table
```

Format

A list contain different types like species, tissue_class, tissue_type, cancer_type, cell_type

Details

This list is used to choose filters for creation of standardized marker list.

Source

http://117.50.127.228/CellMarker/

See Also

Other SlimR_Database: Cellmarker2, Cellmarker2_raw, Markers_list_TCellSI, Markers_list_scIBD, PanglaoDB, PanglaoDB_raw, PanglaoDB_table

---

Celltype_Annotation *Annotate Seurat Object with SlimR Cell Type Predictions*

Description

This function assigns SlimR predicted cell types to a Seurat object based on cluster annotations, and stores the results in the meta.data slot.

Usage

```
Celltype_Annotation(
  seurat_obj,
  cluster_col,
  SlimR_anno_result,
  plot_UMAP = TRUE,
  annotation_col = "Cell_type_SlimR"
)
```

Arguments

seurat_obj	A Seurat object containing cluster information in meta.data.
cluster_col	Character string indicating the column name in meta.data that contains cluster IDs.
SlimR_anno_result	List generated by function Celltype_Calculate() which containing a data.frame in $Prediction_results with: 1.cluster_col (Cluster identifiers (should match cluster_col in meta.data)) 2.Predicted_cell_type (Predicted cell types for each cluster).
plot_UMAP	logical(1); if TRUE, plot the UMAP with cell type annotations.
annotation_col	The location to write in 'meta.data' that contains the predicted cell type. (default = "Cell_type_SlimR")

Value

A Seurat object with updated meta.data containing the predicted cell types.

Note

If plot_UMAP = TRUE, this function will print a UMAP plot as a side effect.

See Also

Other Automated_Annotation_Workflow: `Celltype_Calculate()`, `Celltype_Verification()`

Examples

```
## Not run:
sce <- Celltype_Annotation(seurat_obj = sce,
    cluster_col = "seurat_clusters",
    SlimR_anno_result = SlimR_anno_result,
    plot_UMAP = TRUE,
    annotation_col = "Cell_type_SlimR"
    )

## End(Not run)
```

`Celltype_annotation_Cellmarker2` *Uses "marker_list" from Cellmarker2 for cell annotation*

Description

Uses "marker_list" from Cellmarker2 for cell annotation

Usage

```
Celltype_annotation_Cellmarker2(
  seurat_obj,
  gene_list,
  species,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  save_path = NULL,
  min_counts = 1,
  colour_low = "white",
  colour_high = "navy",
  colour_low_mertic = "white",
  colour_high_mertic = "navy"
)
```

Arguments

seurat_obj	Enter the Seurat object with annotation columns such as "seurat_cluster" in meta.data to be annotated.
gene_list	Enter the standard "Marker_list" generated by the Cellmarker2 database for the SlimR package, generated by the "Markers_filter_Cellmarker2 ()" function.
species	This parameter selects the species "Human" or "Mouse" for standard gene format correction of markers entered by "Marker_list".
cluster_col	Enter annotation columns such as "seurat_cluster" in meta.data of the Seurat object to be annotated. Default parameters use "cluster_col = 'seurat_clusters'".
assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = "RNA"".
save_path	The output path of the cell annotation picture. Example parameters use "save_path = './SlimR/Celltype_annotation_Cellmarker2/'".
min_counts	The minimum number of counts of genes in "Marker_list" entered. This number represents the number of the same gene in the same species and the same location in the Cellmarker2 database used for annotation of this cell type. Default parameters use "min_counts = 1".
colour_low	Color for lowest expression level. (default = "white")
colour_high	Color for highest expression level. (default = "black")
colour_low_mertic	Color for lowest mertic level. (default = "white")
colour_high_mertic	Color for highest mertic level. (default = "black")

Value

The cell annotation picture is saved in "save_path".

See Also

Other Other_Functions_Provided_By_SlimR: Celltype_annotation_Excel(), Celltype_annotation_PanglaoDB(), Celltype_annotation_Seurat()

Examples

```
## Not run:
Celltype_annotation_Cellmarker2(seurat_obj = sce,
    gene_list = Markers_list_Cellmarker2,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_Cellmarker2")
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )
```

```
## End(Not run)
```

`Celltype_Annotation_Combined` *Uses "marker_list" to generate combined plot for cell annotation*

Description

Uses "marker_list" to generate combined plot for cell annotation

Usage

```
Celltype_Annotation_Combined(
  seurat_obj,
  gene_list,
  species,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  save_path = NULL,
  colour_low = "white",
  colour_high = "navy"
)
```

Arguments

`seurat_obj`	Enter the Seurat object with annotation columns such as "seurat_cluster" in meta.data to be annotated.
`gene_list`	A list of cells and corresponding gene controls, the name of the list is cell type, and the first column of the list corresponds to markers. Lists can be generated using functions such as "Markers_filter_Cellmarker2 ()", "Markers_filter_PanglaoDB ()", "read_excel_markers ()", "read_seurat_markers ()", etc.
`species`	This parameter selects the species "Human" or "Mouse" for standard gene format correction of markers entered by "Marker_list".
`cluster_col`	Enter annotation columns such as "seurat_cluster" in meta.data of the Seurat object to be annotated. Default parameters use "cluster_col = 'seurat_clusters'".
`assay`	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = 'RNA'".
`save_path`	The output path of the cell annotation picture. Example parameters use "save_path = './SlimR/Celltype_annotation_Bar/'".
`colour_low`	Color for lowest expression level. (default = "white")
`colour_high`	Color for highest expression level. (default = "black")

Value

The cell annotation picture is saved in "save_path".

See Also

Other Semi_Automated_Annotation_Workflow: `Celltype_Annotation_Features()`, `Celltype_Annotation_Heatmap()`

Examples

```
## Not run:
Celltype_Annotation_Combined(seurat_obj = sce,
    gene_list = Markers_list,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_Annotation_Combined"),
    colour_low = "white",
    colour_high = "navy"
    )

## End(Not run)
```

Celltype_annotation_Excel

Uses "marker_list" from Excel input for cell annotation

Description

Uses "marker_list" from Excel input for cell annotation

Usage

```
Celltype_annotation_Excel(
  seurat_obj,
  gene_list,
  species,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  save_path = NULL,
  metric_names = NULL,
  colour_low = "white",
  colour_high = "navy",
  colour_low_mertic = "white",
  colour_high_mertic = "navy"
)
```

Arguments

seurat_obj	Enter the Seurat object with annotation columns such as "seurat_cluster" in meta.data to be annotated.
gene_list	Enter the standard "Marker_list" generated by the Excel files database for the SlimR package, generated by the "read_excel_markers()" function.
species	This parameter selects the species "Human" or "Mouse" for standard gene format correction of markers entered by "Marker_list".
cluster_col	Enter annotation columns such as "seurat_cluster" in meta.data of the Seurat object to be annotated. Default parameters use "cluster_col = "seurat_clusters"".
assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = 'RNA'".
save_path	The output path of the cell annotation picture. Example parameters use "save_path = './SlimR/Celltype_annotation_Excel/'".
metric_names	Change the row name for the input mertics, not recommended unless necessary. (NULL is used as default parameter)
colour_low	Color for lowest expression level. (default = "white")
colour_high	Color for highest expression level. (default = "black")
colour_low_mertic	Color for lowest mertic level. (default = "white")
colour_high_mertic	Color for highest mertic level. (default = "black")

Value

The cell annotation picture is saved in "save_path".

See Also

Other Other_Functions_Provided_By_SlimR: Celltype_annotation_Cellmarker2(), Celltype_annotation_PanglaoDB
Celltype_annotation_Seurat()

Examples

```
## Not run:
Celltype_annotation_Excel(seurat_obj = sce,
    gene_list = Markers_list_Excel,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_Excel")
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )

## End(Not run)
```

`Celltype_Annotation_Features` *Annotate cell types using features plot with different marker databases*

Description

This function dynamically selects the appropriate annotation method based on the `gene_list_type` parameter. It supports marker databases from Cellmarker2, PanglaoDB, Seurat (via `FindAllMarkers`), or Excel files.

Usage

```
Celltype_Annotation_Features(
  seurat_obj,
  gene_list,
  gene_list_type = "Default",
  species = NULL,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  save_path = NULL,
  min_counts = 1,
  metric_names = NULL,
  colour_low = "white",
  colour_high = "navy",
  colour_low_mertic = "white",
  colour_high_mertic = "navy",
  ...
)
```

Arguments

`seurat_obj`	A valid Seurat object with cluster annotations in `meta.data`.
`gene_list`	A list of data frames containing marker genes and metrics. Format depends on `gene_list_type`: - **Cellmarker2**: Generated by `Markers_filter_Cellmarker2()`. - **PanglaoDB**: Generated by `Markers_filter_PanglaoDB()`. - **Seurat**: Generated by `read_seurat_markers()`. - **Excel**: Generated by `read_excel_markers()`.
`gene_list_type`	Type of marker database to use. Be one of: `"Cellmarker2"`, `"PanglaoDB"`, `"Seurat"`, or `"Excel"`.
`species`	Species of the dataset: `"Human"` or `"Mouse"` for gene name standardization.
`cluster_col`	Column name in `meta.data` defining clusters (default: `"seurat_clusters"`).
`assay`	Assay layer in the Seurat object (default: `"RNA"`).
`save_path`	Directory to save output PNGs. Must be explicitly specified.
`min_counts`	Minimum number of counts for Cellmarker2 annotations (default: 1).
`metric_names`	Optional. Change the row name for the input mertics, not recommended unless necessary. (NULL is used as default parameter; used in "Seurat"/"Excel").

colour_low	Color for lowest expression level. (default = "white")
colour_high	Color for highest expression level. (default = "black")
colour_low_mertic	Color for lowest mertic level. (default = "white")
colour_high_mertic	Color for highest mertic level. (default = "black")
...	Additional parameters passed to the specific annotation function.

Value

Saves cell type annotation PNGs in save_path. Returns invisibly.

See Also

Other Semi_Automated_Annotation_Workflow: Celltype_Annotation_Combined(), Celltype_Annotation_Heatmap()

Examples

```
## Not run:
# Example for Cellmarker2
Celltype_Annotation_Features(seurat_obj = sce,
    gene_list = Markers_list_Cellmarker2,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_Cellmarker2"),
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )

# Example for PanglaoDB
Celltype_Annotation_Features(seurat_obj = sce,
    gene_list = Markers_list_panglaoDB,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_PanglaoDB")
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )

# Example for Seurat marker list
Celltype_Annotation_Features(seurat_obj = sce,
    gene_list = Markers_list_Seurat,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
```

```
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_Seurat")
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )

# Example for Excel marker list
Celltype_Annotation_Features(seurat_obj = sce,
    gene_list = Markers_list_Excel,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_Excel")
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )

## End(Not run)
```

`Celltype_Annotation_Heatmap` *Uses "marker_list" to generate heatmap for cell annotation*

Description

Uses "marker_list" to generate heatmap for cell annotation

Usage

```
Celltype_Annotation_Heatmap(
  seurat_obj,
  gene_list,
  species,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  min_expression = 0.1,
  specificity_weight = 3,
  colour_low = "navy",
  colour_high = "firebrick3"
)
```

Arguments

`seurat_obj`	Enter the Seurat object with annotation columns such as "seurat_cluster" in meta.data to be annotated.

gene_list	A list of cells and corresponding gene controls, the name of the list is cell type, and the first column of the list corresponds to markers. Lists can be generated using functions such as "Markers_filter_Cellmarker2 ()", "Markers_filter_PanglaoDB ()", "read_excel_markers ()", "read_seurat_markers ()", etc.
species	This parameter selects the species "Human" or "Mouse" for standard gene format correction of markers entered by "Marker_list".
cluster_col	Enter annotation columns such as "seurat_cluster" in meta.data of the Seurat object to be annotated. Default parameters use "cluster_col = 'seurat_clusters'".
assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = 'RNA'".
min_expression	The min_expression parameter defines a threshold value to determine whether a cell's expression of a feature is considered "expressed" or not. It is used to filter out low-expression cells that may contribute noise to the analysis. Default parameters use "min_expression = 0.1".
specificity_weight	The specificity_weight parameter controls how much the expression variability (standard deviation) of a feature within a cluster contributes to its "specificity score." It amplifies or suppresses the impact of variability in the final score calculation.Default parameters use "specificity_weight = 3".
colour_low	Color for lowest probability level in Heatmap visualization of probability matrix. (default = "navy")
colour_high	Color for highest probability level Heatmap visualization of probability matrix. (default = "firebrick3")

Value

The heatmap of the comparison between "cluster_col" in the Seurat object and the given gene set "gene_list" needs to be annotated.

See Also

Other Semi_Automated_Annotation_Workflow: `Celltype_Annotation_Combined()`, `Celltype_Annotation_Features()`

Examples

```
## Not run:
Celltype_Annotation_Heatmap(seurat_obj = sce,
    gene_list = Markers_list,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    min_expression = 0.1,
    specificity_weight = 3,
    colour_low = "navy",
    colour_high = "firebrick3"
    )

## End(Not run)
```

`Celltype_annotation_PanglaoDB` *Uses "marker_list" from PanglaoDB for cell annotation*

Description

Uses "marker_list" from PanglaoDB for cell annotation

Usage

```
Celltype_annotation_PanglaoDB(
  seurat_obj,
  gene_list,
  species,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  save_path = NULL,
  metric_names = NULL,
  colour_low = "white",
  colour_high = "navy",
  colour_low_mertic = "white",
  colour_high_mertic = "navy"
)
```

Arguments

`seurat_obj`	Enter the Seurat object with annotation columns such as "seurat_cluster" in meta.data to be annotated.
`gene_list`	Enter the standard "Marker_list" generated by the PanglaoDB database for the SlimR package, generated by the "Markers_filter_PanglaoDB ()" function.
`species`	This parameter selects the species "Human" or "Mouse" for standard gene format correction of markers entered by "Marker_list".
`cluster_col`	Enter annotation columns such as "seurat_cluster" in meta.data of the Seurat object to be annotated. Default parameters use "cluster_col = 'seurat_clusters'".
`assay`	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = 'RNA'".
`save_path`	The output path of the cell annotation picture. Example parameters use "save_path = './SlimR/Celltype_annotation_PanglaoDB/'".
`metric_names`	Warning: Do not enter information. This parameter is used to check if "Marker_list" conforms to the PanglaoDB database output.
`colour_low`	Color for lowest expression level. (default = "white")
`colour_high`	Color for highest expression level. (default = "black")
`colour_low_mertic`	Color for lowest mertic level. (default = "white")
`colour_high_mertic`	Color for highest mertic level. (default = "black")

Value

The cell annotation picture is saved in "save_path".

See Also

Other Other_Functions_Provided_By_SlimR: Celltype_annotation_Cellmarker2(), Celltype_annotation_Excel(), Celltype_annotation_Seurat()

Examples

```
## Not run:
Celltype_annotation_PanglaoDB(seurat_obj = sce,
    gene_list = Markers_list_panglaoDB,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_PanglaoDB")
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )

## End(Not run)
```

Celltype_annotation_Seurat

Uses "marker_list" from Seurat object for cell annotation

Description

Uses "marker_list" from Seurat object for cell annotation

Usage

```
Celltype_annotation_Seurat(
  seurat_obj,
  gene_list,
  species,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  save_path = NULL,
  metric_names = NULL,
  colour_low = "white",
  colour_high = "navy",
  colour_low_mertic = "white",
  colour_high_mertic = "navy"
)
```

Arguments

seurat_obj	Enter the Seurat object with annotation columns such as "seurat_cluster" in meta.data to be annotated.
gene_list	Enter the standard "Marker_list" generated by the Seurat object database for the SlimR package, generated by the "read_seurat_markers()" function.
species	This parameter selects the species "Human" or "Mouse" for standard gene format correction of markers entered by "Marker_list".
cluster_col	Enter annotation columns such as "seurat_cluster" in meta.data of the Seurat object to be annotated. Default parameters use "cluster_col = 'seurat_clusters'".
assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = 'RNA'".
save_path	The output path of the cell annotation picture. Example parameters use "save_path = './SlimR/Celltype_annotation_Seurat/'".
metric_names	Change the row name for the input mertics, not recommended unless necessary. (NULL is used as default parameter)
colour_low	Color for lowest expression level. (default = "white")
colour_high	Color for highest expression level. (default = "black")
colour_low_mertic	Color for lowest mertic level. (default = "white")
colour_high_mertic	Color for highest mertic level. (default = "black")

Value

The cell annotation picture is saved in "save_path".

See Also

Other Other_Functions_Provided_By_SlimR: Celltype_annotation_Cellmarker2(), Celltype_annotation_Excel(), Celltype_annotation_PanglaoDB()

Examples

```
## Not run:
Celltype_annotation_Seurat(seurat_obj = sce,
    gene_list = Markers_list_Seurat,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    save_path = file.path(tempdir(),"SlimR_Celltype_annotation_Seurat")
    colour_low = "white",
    colour_high = "navy",
    colour_low_mertic = "white",
    colour_high_mertic = "navy",
    )

## End(Not run)
```

Celltype_Calculate *Uses "marker_list" to calculate probability, prediction results, AUC and generate heatmap for cell annotation*

Description

Uses "marker_list" to calculate probability, prediction results, AUC and generate heatmap for cell annotation

Usage

```
Celltype_Calculate(
  seurat_obj,
  gene_list,
  species,
  cluster_col = "seurat_clusters",
  assay = "RNA",
  min_expression = 0.1,
  specificity_weight = 3,
  threshold = 0.8,
  compute_AUC = TRUE,
  plot_AUC = TRUE,
  AUC_correction = TRUE,
  colour_low = "navy",
  colour_high = "firebrick3"
)
```

Arguments

seurat_obj	Enter the Seurat object with annotation columns such as "seurat_cluster" in meta.data to be annotated.
gene_list	A list of cells and corresponding gene controls, the name of the list is cell type, and the first column of the list corresponds to markers. Lists can be generated using functions such as "Markers_filter_Cellmarker2 ()", "Markers_filter_PanglaoDB ()", "read_excel_markers ()", "read_seurat_markers ()", etc.
species	This parameter selects the species "Human" or "Mouse" for standard gene format correction of markers entered by "Marker_list".
cluster_col	Enter annotation columns such as "seurat_cluster" in meta.data of the Seurat object to be annotated. Default parameters use "cluster_col = 'seurat_clusters'".
assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = 'RNA'".
min_expression	The min_expression parameter defines a threshold value to determine whether a cell's expression of a feature is considered "expressed" or not. It is used to filter out low-expression cells that may contribute noise to the analysis. Default parameters use "min_expression = 0.1".

`specificity_weight`
: The specificity_weight parameter controls how much the expression variability (standard deviation) of a feature within a cluster contributes to its "specificity score." It amplifies or suppresses the impact of variability in the final score calculation.Default parameters use "specificity_weight = 3".

`threshold`
: This parameter refers to the normalized similarity between the "alternative cell type" and the "predicted cell type" in the returned results. (the default parameter is 0.8)

`compute_AUC`
: Logical indicating whether to calculate AUC values for predicted cell types. AUC measures how well the marker genes distinguish the cluster from others. When TRUE, adds an AUC column to the prediction results. (default: TRUE)

`plot_AUC`
: The logic indicates whether to draw an AUC curve for the predicted cell type. When TRUE, add an AUC_plot to result. (default: TRUE)

`AUC_correction`
: Logical value controlling AUC-based correction. (default = TRUE) When set to TRUE: 1.Computes AUC values for candidate cell types. (probability > threshold) 2.Selects the cell type with the highest AUC as the final predicted type. 3.Records the selected type's AUC value in the "AUC" column.

`colour_low`
: Color for lowest probability level in Heatmap visualization of probability matrix. (default = "navy")

`colour_high`
: Color for highest probability level Heatmap visualization of probability matrix. (default = "firebrick3")

Value

A list containing:

- Expression_list: List of expression matrices for each cell type
- Proportion_list: List of proportion of expression for each cell type
- Expression_scores_matrix: Matrix of expression scores
- Probability_matrix: Matrix of normalized probabilities
- Prediction_results: Data frame with cluster annotations including:
  - cluster_col: Cluster identifier
  - Predicted_cell_type: Primary predicted cell type
  - AUC: Area Under the Curve value (when compute_AUC = TRUE)
  - Alternative_cell_types: Semi-colon separated alternative cell types
- Heatmap_plot: Heatmap visualization of probability matrix
- AUC_plot: AUC visualization of Predicted cell type

See Also

Other Automated_Annotation_Workflow: `Celltype_Annotation()`, `Celltype_Verification()`

Examples

```
## Not run:
SlimR_anno_result <- Celltype_Calculate(seurat_obj = sce,
    gene_list = Markers_list,
    species = "Human",
    cluster_col = "seurat_clusters",
    assay = "RNA",
    min_expression = 0.1,
    specificity_weight = 3,
    threshold = 0.8,
    compute_AUC = TRUE,
    plot_AUC = TRUE,
    AUC_correction = TRUE,
    colour_low = "navy",
    colour_high = "firebrick3"
    )

## End(Not run)
```

Celltype_Verification *Perform cell type verification and generate the validation dotplot*

Description

This function performs verification of predicted cell types by selecting high log2FC and high expression proportion genes and generates and generate the validation dotplot.

Usage

```
Celltype_Verification(
  seurat_obj,
  SlimR_anno_result,
  assay = "RNA",
  gene_number = 5,
  colour_low = "white",
  colour_high = "navy",
  annotation_col = "Cell_type_SlimR"
)
```

Arguments

seurat_obj	A Seurat object containing single-cell data.
SlimR_anno_result	A list containing SlimR annotation results with: Expression_list - List of expression matrices for each cell type. Prediction_results - Data frame with cluster annotations.

assay	Enter the assay used by the Seurat object, such as "RNA". Default parameters use "assay = 'RNA'".
gene_number	Integer specifying number of top genes to select per cell type.
colour_low	Color for lowest expression level. (default = "white")
colour_high	Color for highest expression level. (default = "black")
annotation_col	Character string specifying the column in meta.data to use for grouping.

Value

A ggplot object showing expression of top variable genes.

See Also

Other Automated_Annotation_Workflow: `Celltype_Annotation()`, `Celltype_Calculate()`

Examples

```
## Not run:
Celltype_Verification(seurat_obj = sce,
    SlimR_anno_result = SlimR_anno_result,
    assay = "RNA",
    gene_number = 5,
    colour_low = "white",
    colour_high = "navy",
    annotation_col = "Cell_type_SlimR"
    )

## End(Not run)
```

Markers_filter_Cellmarker2

Create Marker_list from the Cellmarkers2 database

Description

Create Marker_list from the Cellmarkers2 database

Usage

```
Markers_filter_Cellmarker2(
  df,
  species = NULL,
  tissue_class = NULL,
  tissue_type = NULL,
  cancer_type = NULL,
  cell_type = NULL
)
```

Arguments

df	Standardized Cellmarkers2 database. It is read as data(Cellmarkers2) in the SlimR library.
species	Species information in Cellmarkers2 database. The default input is "Human" or "Mouse".The input can be retrieved by "Cellmarkers2_table". For more information,please refer to http://117.50.127.228/CellMarker/ on Cellmarkers2's official website.
tissue_class	Tissue_class information in Cellmarkers2 database. The input can be retrieved by "Cellmarkers2_table". For more information, please refer to http://117.50.127.228/CellMarker/ on Cellmarkers2's official website.
tissue_type	Tissue_type information in Cellmarkers2 database. The input can be retrieved by "Cellmarkers2_table". For more information, please refer to http://117.50.127.228/CellMarker/ on Cellmarkers2's official website.
cancer_type	Cancer_type information in Cellmarkers2 database. The input can be retrieved by "Cellmarkers2_table". For more information, please refer to http://117.50.127.228/CellMarker/ on Cellmarkers2's official website.
cell_type	Cell_type information in Cellmarkers2 database. The input can be retrieved by "Cellmarkers2_table". For more information, please refer to http://117.50.127.228/CellMarker/ on Cellmarkers2's official website.

Value

The standardized "Marker_list" in the SlimR package

See Also

Other Standardized_Marker_list_Input: `Markers_filter_PanglaoDB()`, `Read_excel_markers()`, `Read_seurat_markers()`

Examples

```
Cellmarker2 <- SlimR::Cellmarker2
Markers_list_Cellmarker2 <- Markers_filter_Cellmarker2(
    Cellmarker2,
    species = "Human",
    tissue_class = "Intestine",
    tissue_type = NULL,
    cancer_type = NULL,
    cell_type = NULL
    )
```

`Markers_filter_PanglaoDB` *Create Marker_list from the PanglaoDB database*

Description

Create Marker_list from the PanglaoDB database

Usage

```
Markers_filter_PanglaoDB(df, species_input, organ_input)
```

Arguments

`df`	Standardized PanglaoDB database. It is read as data(PanglaoDB) in the SlimR library.
`species_input`	Species information in PanglaoDB database. The default input is "Human" or "Mouse".The input can be retrieved by "PanglaoDB_table". For more information,please refer to https://panglaodb.se/ on PanglaoDB's official website.
`organ_input`	Organ type information in the PanglaoDB database. The input can be retrieved by "PanglaoDB_table".For more information, please refer to https://panglaodb.se/ on PanglaoDB's official website.

Value

The standardized "Marker_list" in the SlimR package

See Also

Other Standardized_Marker_list_Input: `Markers_filter_Cellmarker2()`, `Read_excel_markers()`, `Read_seurat_markers()`

Examples

```
PanglaoDB <- SlimR::PanglaoDB
Markers_list_panglaoDB <- Markers_filter_PanglaoDB(
    PanglaoDB,
    species_input = 'Human',
    organ_input = 'GI tract'
    )
```

Markers_list_scIBD *List of cell type markers in the scIBD dataset*

Description

A dataset containing marker genes for different human intestine cell types from scIBD

Usage

```
Markers_list_scIBD
```

Format

A list with one hundred and one tables.

Details

This list is a table of 101 types of human intestine cell types markers obtained from scIBD. The article doi source is "https://doi.org/10.1038/s43588-023-00464-9", and the reference literature is: Nie et al. (2023) doi:10.1038/s43588-023-00464-9.

Source

doi:10.1038/s43588023004649

See Also

Other SlimR_Database: Cellmarker2, Cellmarker2_raw, Cellmarker2_table, Markers_list_TCellSI, PanglaoDB, PanglaoDB_raw, PanglaoDB_table

Markers_list_TCellSI *List of cell type markers in the TCellSI dataset*

Description

A dataset containing marker genes for different T cell types from TCellSI

Usage

```
Markers_list_TCellSI
```

Format

A list with ten tables.

Details

This list is a table of 10 types of T cell markers obtained from TCellSI. The data source is "https://github.com/GuoBioinfoLab/ and the reference literature is: Yang et al. (2024) doi:10.1002/imt2.231.

Source

https://github.com/GuoBioinfoLab/TCellSI/

See Also

Other SlimR_Database: Cellmarker2, Cellmarker2_raw, Cellmarker2_table, Markers_list_scIBD, PanglaoDB, PanglaoDB_raw, PanglaoDB_table

---

PanglaoDB *PanglaoDB dataset*

Description

A dataset containing marker genes for different cell types from PanglaoDB

Usage

```
PanglaoDB
```

Format

A data frame with 9 columns:

Details

This dataset is used to filter and create a standardized marker list.'

Source

https://panglaodb.se/

See Also

Other SlimR_Database: Cellmarker2, Cellmarker2_raw, Cellmarker2_table, Markers_list_TCellSI, Markers_list_scIBD, PanglaoDB_raw, PanglaoDB_table

PanglaoDB_raw *PanglaoDB raw dataset*

Description

A dataset containing marker genes for different cell types from PanglaoDB

Usage

```
PanglaoDB_raw
```

Format

A data frame with 14 columns contined in the PanglaoDB database:

Details

This dataset is used to filter and create a standardized marker list.'

Source

https://panglaodb.se/

See Also

Other SlimR_Database: Cellmarker2, Cellmarker2_raw, Cellmarker2_table, Markers_list_TCellSI, Markers_list_scIBD, PanglaoDB, PanglaoDB_table

PanglaoDB_table *PanglaoDB table*

Description

A dataset containing marker genes for different cell types from PanglaoDB

Usage

```
PanglaoDB_table
```

Format

A list contain different types like species, organ, cell type.

Details

This list is used to choose filters for creation of standardized marker list.

Source

https://panglaodb.se/

See Also

Other SlimR_Database: Cellmarker2, Cellmarker2_raw, Cellmarker2_table, Markers_list_TCellSI, Markers_list_scIBD, PanglaoDB, PanglaoDB_raw

---

Read_excel_markers *Create "Marker_list" from Excel files ".xlsx"*

Description

Create "Marker_list" from Excel files ".xlsx"

Usage

```
Read_excel_markers(path)
```

Arguments

path	The path information of Marker files stored in ".xlsx" format. The Sheet name in the file is filled with cell type. The first line of each Sheet is the table head, the first column is filled with markers information, and the following column is filled with mertic information.

Value

The standardized "Marker_list" in the SlimR package.

See Also

Other Standardized_Marker_list_Input: Markers_filter_Cellmarker2(), Markers_filter_PanglaoDB(), Read_seurat_markers()

Examples

```
## Not run:
Markers_list_Excel <- Read_excel_markers(
    "D:/Laboratory/Marker_load.xlsx"
    )

## End(Not run)
```

Read_seurat_markers *Create "Marker_list" from Seurat object*

Description

Create "Marker_list" from Seurat object

Usage

```
Read_seurat_markers(
  df,
  sources = c("Seurat", "presto"),
  sort_by = "FSS",
  gene_filter = 20
)
```

Arguments

df	Dataframe generated by "FindAllMarkers" function, recommend to use parameter "group.by = "Cell_type"" and "only.pos = TRUE".
sources	Type of markers sources to use. Be one of: "Seurat" or "presto".
sort_by	Marker sorting parameter, select "avg_log2FC" or "p_val_adj" or "FSS" (Feature Significance Score, FSS, product value of log2FC and Expression ratio). Default parameters use "sort_by = 'FSS'".
gene_filter	The number of markers left for each cell type based on the "sort_by" parameter's level of difference. Default parameters use "gene_fliter = 20"

Value

The standardized "Marker_list" in the SlimR package.

See Also

Other Standardized_Marker_list_Input: Markers_filter_Cellmarker2(), Markers_filter_PanglaoDB(), Read_excel_markers()

Examples

```
## Not run:
# Example for Seurat sources markers
seurat_markers <- Seurat::FindAllMarkers(
    object = sce,
    group.by = "Cell_type",
    only.pos = TRUE)

Markers_list_Seurat <- Read_seurat_markers(seurat_markers,
    sources = "Seurat",
```

```
    sort_by = "avg_log2FC",
    gene_filter = 20
    )

# Example for presto sources markers
seurat_markers <- dplyr::filter(
    presto::wilcoxauc(
      X = sce,
      group_by = "Cell_type",
      seurat_assay = "RNA"
      ),
    padj < 0.05, logFC > 0.5
    )

Markers_list_Seurat <- Read_seurat_markers(seurat_markers,
    sources = "presto",
    sort_by = "logFC",
    gene_filter = 20
    )

## End(Not run)
```

Index