Package 'ClusterMap'

June 25, 2018

Type Package

Title What the Package Does (Title Case)

Version 0.1.0

Author Xin Gao

Maintainer Xin Gao <xig@stowers.org>

Description
ClusterMap is designed to analyze and compare two or more single cell expression datasets.

License MIT + file LICENSE

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

Depends R(>= 3.4.3)

Imports Seurat(>= 2.2.1),pheatmap(>= 1.0.10),ape(>= 5.1),circlize(>= 0.4.3)

R topics documented:

---

add_perc   *circos_map*

Description

Plot Circos plot for the matching results.

Usage

```
add_perc(mapRes, cell_num_list)
```

Arguments

mapRes   A dataframe of the output of function cluster_map_by_marker.

cell_num_list

        A list of vector of cell numbers for each group and each sample.

Value

A dataframe of the matching results with cell percentage column added.

---

circos_map   *circos_map*

Description

Plot Circos plot for the matching results.

Usage

```
circos_map(mapRes, cell_num_list, output, color_cord = NULL,
  color_sample = NULL)
```

Arguments

mapRes   A dataframe of the output of function cluster_map_by_marker.

cell_num_list

        A list of vector of cell numbers for each group and each sample.

output   The output directory to save the plot.

color_cord   A vector of colors for the cord of circos plot. DEFAULT is NULL. Pre defined internal color will be used.

color_sample   A vector of colors for the sample sectors in the circos plot. DEFAULT is NULL. Pre defined internal color will be used.

Value

circos plot will be save.

`cluster_map`   *cluster_map*

Description

A master function to perform the full workflow of ClusterMap.

Usage

```
cluster_map(marker_file_list, edge_cutoff = 0.1, output,
  cell_num_list = NULL, single_obj_list = NULL, comb_obj = NULL,
  comb_delim = "-", k = 5)
```

Arguments

`marker_file_list`

A list of csv files with names. Each file is a marker gene table for a sample. The columns named as 'cluster' and 'gene' are required.

`edge_cutoff`   The edge length cutoff to decide the sub-nodes to merge or not. DEFAULT is 0.1.

`output`   The output directory to save the matching results.

`cell_num_list`

A list of vector of cell numbers for each group and each sample.

`single_obj_list`

A list of Seurat object for each sample, with the same list names as the list names of marker_file_list.

`comb_obj`   A Seurat object for the combined sample. Cells in different samples are labelled by the sample names with the comb_delim. The sample names should be the same as the list names of marker_file_list.

`comb_delim`   The delimiter used in the cell names in the combined object to connect sample name and cell name in individual sample. DEFAULT is '-'.

`k`   K-nearest neighbours used to calculate distance. DEFAULT is 5.

Value

A dataframe of the matching results. Heatmap of marker genes, the corresponding dendrogram, circos plot and recolored t-SNE plots will be saved into files.

`cluster_map_by_marker`
                    cluster_map_by_marker

Description

Match groups by marker genes and decompose into new groups by purity cut.

Usage

```
cluster_map_by_marker(marker_file_list, cutoff = 0.1, output)
```

Arguments

marker_file_list

A list of csv files. Each file is a marker gene table for a sample. The columns named as 'cluster' and 'gene' are required.

cutoff   The edge length cutoff to decide the sub-nodes to merge or not. DEFAULT is 0.1.

output   The output directory to save the matching results.

Value

A dataframe of the matching results. Heatmap of marker genes and the dendrogram will be saved into files.

---

inna_dist   *inna_dist*

Description

Calculate the inna-sample distance. Internal function called by separability.

Usage

```
inna_dist(x, k)
```

Arguments

x   A distance matrix of cells in one sample to the cells in the same sample.

k   K-nearest neighbours used to calculate distance.

Value

A single value of distance.

---

inter_dist *inter_dist*

Description

Calculate the inter-sample distance. Internal function called by separability.

Usage

```
inter_dist(x, k)
```

Arguments

x	A distance matrix of cells in one sample to the cells in another sample.
k	K-nearest neighbours used to calculate distance. DEFAULT is 5.

Value

A single value of distance.

---

make_comb_obj *make_comb_obj*

Description

A warper of Seurat function to generate Seurat object for combined sample from single sample.

Usage

```
make_comb_obj(data_dirList, is.10X = TRUE, comb_delim = "-")
```

Arguments

data_dirList	A list of directory with 10X genomics single cell data or full path of expression table.
is.10X	The input data is 10X genomics format or not. DEFAULT is TRUE.
comb_delim	The delimiter used in the cell names in the combined object to connect sample name and cell name in individual sample. DEFAULT is '-'.

Value

A Seurat object.

---

make_single_obj *make_single_obj*

Description

A warper of Seurat function to generate Seurat object and marker genes for single sample.

Usage

```
make_single_obj(data_dir, is.10X = TRUE, output)
```

Arguments

data_dir	Directory with 10X genomics single cell data or full path of expression table.
is.10X	The input data is 10X genomics format or not. DEFAULT is TRUE.
output	The output directory to save the marker genes.

Value

A Seurat object.

---

purity_cut *purity_cut*

Description

Cut hierarchical clustering dendrogram by edge length and purity of the nodes.

Usage

```
purity_cut(hcluster, cutoff = 0.1)
```

Arguments

hcluster	A hclust object.
cutoff	The edge length cutoff to decide the sub-nodes to merge or not. DEFAULT is 0.1.

Value

A dataframe of the matching results.

recolor_comb	*recolor_comb*

Description

Recolor the combined sample based on the matching results from recolor_s.

Usage

```
recolor_comb(comb_obj, new_group_list, output, comb_delim = "-",
  color = NULL)
```

Arguments

comb_obj	A Seurat object for the combined sample. Cells in different samples are labelled by the sample names with the comb_delim.
new_group_list	
	A list of vectors of new group assignment outputted from recolor_s.
output	The output directory to save the plot.
comb_delim	The delimiter used in the cell names in the combined object to connect sample name and cell name in individual sample. DEFAULT is '-'.
color	A vector of colors used to recolor the new groups. DEFAULT is NULL. Pre defined internal color will be used.

Value

A vector of new group labels with the cell name as the vector name.

recolor_s	*recolor_s*

Description

Recolor a single sample based on the matching results from cluster_map_by_marker.

Usage

```
recolor_s(mapRes_sub, obj, output, color = NULL)
```

Arguments

mapRes_sub	A vector of the column named by the sample in the output of cluster_map_by_marker, with the regroup column as the vector name.
obj	A Seurat object for the sample.
output	The output directory to save the plot.
color	A vector of colors used to recolor the new groups. DEFAULT is NULL. Pre defined internal color will be used.

Value

A vector of new group labels with the cell name as the vector name.

---

separability *separability*

Description

Calculate separability for labeled data. Internal function called by separability_by_group.

Usage

```
separability(coord, class_label, k = 5)
```

Arguments

coord	A dataframe of the two dimension t-SNE coordinates of each cell in the combined sample.
class_label	A vector of sample labels for each cell, with the same order as the tsne_coord.
k	K-nearest neighbours used to calculate distance. DEFAULT is 5.

Value

A single value of separability.

---

separability_by_group

separability_by_group

Description

Calculate separability for each group in a pair of samples. Internal function called by separability_pairwise.

Usage

```
separability_by_group(tsne_coord, group, sample_label, k = 5)
```

Arguments

tsne_coord	A dataframe of the two dimension t-SNE coordinates of each cell in the combined sample.
group	A vector of group assignment for each cell, with the same order as the tsne_coord.
sample_label	A vector of sample labels for each cell, with the same order as the tsne_coord.
k	K-nearest neighbours used to calculate distance. DEFAULT is 5.

Value

A vector of separability for each group.

separability_pairwise

separability_pairwise

Description

Calculate separability for every sample pair. The higher the more separable.

Usage

```
separability_pairwise(tsne_coord, group, sample_label, k = 5)
```

Arguments

tsne_coord   A dataframe of the two dimension t-SNE coordinates of each cell in the combined sample.

group   A vector of group assignment for each cell, with the same order as the tsne_coord.

sample_label   A vector of sample labels for each cell, with the same order as the tsne_coord.

k   K-nearest neighbours used to calculate distance. DEFAULT is 5.

Value

A matrix of separability for each sample pair (column) and each group (row).