```
## Warning: replacing previous import 'dplyr::count' by 'matrixStats::count' when
## loading 'MOSim'
```

Introduction

MOSim is an R package for the multi-omic simulation of RNA-seq, miRNA-seq, DNase-seq, ChIP-seq and Methyl-seq. In scMOSim, the simulation of single-cell RNA-seq, single-cell ATAC-seq data, and transcription factor information is simulated. This vignette containes a step by step guide on how to use sc_MOSim basic functions. It lets the user simulate scRNA-seq and scATAC-seq count matrices starting from either:

- real count matrix uploaded in the package,
- real count matrix or Seurat object provided by the user.

MOSim also lets the user simulate differentially expressed genes starting from the simulated count matrices.

Data preparation

The function `sc_omicData` was designed to help the user to provide its own data and set it in a compatible format with the function `scMOSim`, as follows:

```
sc_omicData(omics, data = NULL)
```

Real count matrix uploaded in the package

sc_MOSim allows the user to simulate scRNA and scATAC count data resembling the characteristics of PBMC dataset (from a healthy donor) available from 10X Genomics at the following link. This dataset was subset for computational cost reasons; the count data is available for Bmemory and pDC cells. Bmemory cells represent the first 160 cells (columns) of the matrix; pDC cells represent the second 110 cells (columns).

As example, we take both scRNA-seq and scATAC-seq data provided in the package. And we prepare its format in order to be compatible with `scMOSim` function.

```
scRNAseq <- sc_omicData("scRNA-seq")
scATACseq <- sc_omicData("scATAC-seq")

omicsList <- c(scRNAseq, scATACseq)
```

`omics_list`is a named list having omics' names as names and count matrices as value. It will be one of the argument of function `scMOSim`.

Real count matrix or Seurat object provided by the user

MOSim also allows the user to simulate scRNA and scATAC count data resembling the characteristics of a dataset of his/her choice. The supported input formats are :

- count matrix,
- Seurat object.

As example, we load a scRNA-seq count matrix.

```
count_matrix <- example_sc_count_matrix

scRNAseq_user <- sc_omicData("scRNA-seq", count_matrix)
```

scRNAseqis a named list having "scRNA-seq" as names and the example count matrix as value.

Multi-omic simulation

Without providing optional arguments

MOSim gives the user the opportunity to simulate scRNA and scATAC count matrices which resemble the characteristics of the input data, without providing further arguments.

Supposing we're using the data produced in the previous section, we will simulate a scRNAseq and a scATACseq count matrix. As example, the count matrices will have:

- 36601 genes
- 143887 peaks
- 2 cell types A and B
- cell type A has 20 cells
- cell type B has 20 cells

scMOSim, in case of not providing optional arguments, takes as input an omics list and a cell type list.

```
# selecting first 20 cells of Bmemory cells and first 20 cells of pDC cells to
# perform the simulation
cellTypes <- list(cellA = c(1:20), cellB = c(161:191))
sim <- scMOSim(omicsList, cellTypes)
```

```
[1] "Experimental condition 1"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
[1] "Experimental condition 2"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
[1] "Experimental condition 1"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
[1] "Experimental condition 2"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
Number of experimental conditions:  2
Number of genes:  36601
```

```
Number of cells:  51
Setting gene expression intensity...
Setting gene expression variability ...
Simulating biological variability ...
Simulating technical variability ...
Number of experimental conditions:  2
Number of genes:  143887
Number of cells:  51
Setting gene expression intensity...
Setting gene expression variability ...
Simulating biological variability ...
Simulating technical variability ...
```

sim is a named list having 'sim_sc + omic name' as names and Seurat objects as values (where the synthetic count matrices are contained).

Providing optional arguments

In case the user wants to simulate a custom amount of cells per cell type and provide mean and standard deviation for the library size (cell library size), there's the possibility to pass those arguments to `scMOSim`.

As example, the count matrices will have:

- 36601 genes
- 143887 peaks
- 2 cell types A and B
- cell type A has 10 cells, with an average library size of 2M reads
- cell type B has 20 cells, with an average library size of 1M reads

```
sim_with_arg <- scMOSim(omicsList, cellTypes, numberCells = c(10, 20), mean = c(2 *
    10^6, 1 * 10^6), sd = c(10^3, 10^3))
```

```
[1] "Experimental condition 1"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
[1] "Experimental condition 2"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
[1] "Experimental condition 1"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
[1] "Experimental condition 2"
[1] "...estimating gene intensity"
[1] "...estimating gene variability"
[1] "...estimating library size"
[1] "...creating SPARSim simulation parameter"
Number of experimental conditions:  2
```

```
Number of genes:  36601
Number of cells:  30
Setting gene expression intensity...
Setting gene expression variability ...
Simulating biological variability ...
Simulating technical variability ...
Number of experimental conditions:  2
Number of genes:  143887
Number of cells:  30
Setting gene expression intensity...
Setting gene expression variability ...
Simulating biological variability ...
Simulating technical variability ...
```

`sim_with_arg` is a named list having 'sim_sc + omic name' as names and Seurat objects as values (where the synthetic count matrices are contained).