

# Package ‘MOSim’

April 15, 2020

**Title** Multi-Omics Simulation (MOSim)

**Version** 1.0.2

**Description** MOSim package simulates multi-omic experiments that mimic regulatory mechanisms within the cell, allowing flexible experimental design including time course and multiple groups.

**Encoding** UTF-8

**Depends** R (>= 3.6)

**License** GPL-3

**LazyData** false

**biocViews** Software, TimeCourse, ExperimentalDesign, RNASeq

**BugReports** <https://github.com/Neurergus/MOSim/issues>

**URL** <https://github.com/Neurergus/MOSim>

**Imports** HiddenMarkov, zoo, methods, matrixStats, dplyr, stringi, lazyeval, rlang, stats, utils, purrr, scales, stringr, tibble, tidy, ggplot2, Biobase, IRanges, S4Vectors

**Suggests** testthat, knitr, rmarkdown, BiocStyle

**Collate** 'AllClass.R' 'AllGeneric.R' 'Simulator.R' 'SimulatorRegion.R' 'ChIP-seq.R' 'DNase-seq.R' 'functions.R' 'Simulation.R' 'MOSim.R' 'RNA-seq.R' 'simulate\_WGBS\_functions.R' 'methyl-seq.R' 'miRNA-seq.R' 'zzz.R'

**RoxygenNote** 6.1.1

**VignetteBuilder** knitr

**git\_url** <https://git.bioconductor.org/packages/MOSim>

**git\_branch** RELEASE\_3\_10

**git\_last\_commit** 753625b

**git\_last\_commit\_date** 2020-04-03

**Date/Publication** 2020-04-14

**Author** Carlos Martínez [cre, aut],  
Sonia Tarazona [aut]

**Maintainer** Carlos Martínez <cmarmir@gmail.com>

**R topics documented:**

MOSim-package . . . . .	2
experimentalDesign . . . . .	2
is.declared . . . . .	3
mosim . . . . .	3
omicData . . . . .	4
omicResults . . . . .	5
omicSettings . . . . .	6
omicSim . . . . .	7
plotProfile . . . . .	8
sampleData . . . . .	9

**Index** **10**


---

MOSim-package	<i>MOSim</i>
---------------	--------------

---

**Description**

Multiomics simulation package.

---

experimentalDesign	<i>Retrieves the experimental design</i>
--------------------	--

---

**Description**

Retrieves the experimental design

**Usage**

```
experimentalDesign(simulation)
```

**Arguments**

simulation      A MOSimulation object

**Value**

A data frame containing the experimental design used to simulate the data.

**Examples**

```
omic_list <- c("RNA-seq")
rnaseq_simulation <- mosim(omics = omic_list)
# This will be a data frame with RNA-seq counts

design_matrix <- experimentalDesign(rnaseq_simulation)
```

---

is.declared	<i>Check if a variable is declared.</i>
-------------	---

---

**Description**

Check if a variable is declared.

**Usage**

```
is.declared(object, key = NULL)
```

**Arguments**

object	Variable name to check
key	Optional key to check inside object.

**Value**

TRUE or FALSE indicating if the variable is initialized & non-empty.

---

mosim	<i>mosim</i>
-------	--------------

---

**Description**

Performs a multiomic simulation by chaining two actions: 1) Creating the "MOSimulation" class with the provided params. 2) Calling "simulate" method on the initialized object.

**Usage**

```
mosim(omics, omicsOptions, diffGenes, numberReps, numberGroups, times,
      depth, profileProbs, minMaxFC, TFtoGene)
```

**Arguments**

omics	Character vector containing the names of the omics to simulate, which can be "RNA-seq", "miRNA-seq", "DNase-seq", "ChIP-seq" or "Methyl-seq" (e.g. c("RNA-seq", "miRNA-seq")). It can also be a list with the omic names as names and their options as values, but we recommend to use the argument <code>omicSim</code> to provide the options to simulated each omic.
omicsOptions	List containing the options to simulate each omic. We recommend to apply the helper method <code>omicSim</code> to create this list in a friendly way, and the function <code>omicData</code> to provide custom data (see the related sections for more information). Each omic may have different configuration parameters, but the common ones are:  <b>simuData/idToGene</b> Seed sample and association tables for regulatory omics. The helper function <code>omicData</code> should be used to provide this information (see the following section).

	<b>regulatorEffect</b> For regulatory omics. List containing the percentage of effect types (repressor, activator or no effect) over the total number of regulators. See vignette for more information.
	<b>totalFeatures</b> Number of features to simulate. By default, the total number of features in the seed dataset.
	<b>depth</b> Sequencing depth in millions of reads. If not provided, it takes the global parameter passed to <code>mosim</code> function.
	<b>replicateParams</b> List with parameters <i>a</i> and <i>b</i> for adjusting the variability in the generation of replicates using the negative binomial. See vignette for more information.
<code>diffGenes</code>	Number of differentially expressed genes to simulate, given in percentage (0 - 1) or in absolute number (> 1). By default 0.15
<code>numberReps</code>	Number of replicates per experimental condition (and time point, if time series are to be generated). By default 3.
<code>numberGroups</code>	Number of experimental groups or conditions to simulate.
<code>times</code>	Vector of time points to consider in the experimental design.
<code>depth</code>	Sequencing depth in millions of reads.
<code>profileProbs</code>	Numeric vector with the probabilities to assign each of the patterns. Defaults to 0.2 for each.
<code>minMaxFC</code>	Numeric vector of length 2 with minimum and maximum fold-change for differentially expressed features, respectively.
<code>TFtoGene</code>	A logical value indicating if default transcription factors data should be used (TRUE) or not (FALSE), or a 3 column data frame containing custom associations. By default FALSE.

**Value**

Instance of class "MOSimulation" containing the multiomic simulation data.

**Examples**

```

moSimulation <- mosim(
  omics = c("RNA-seq"),
  numberReps = 3,
  times = c(0, 2, 6, 12, 24)
)

# Retrieve simulated count matrix for RNA-seq
dataRNAseq <- omicResults(moSimulation, "RNA-seq")

```

---

omicData

*Set customized data for an omic.*


---

**Description**

Set customized data for an omic.

**Usage**

```
omicData(omic, data = NULL, associationList = NULL)
```

**Arguments**

**omic**                    The name of the omic to provide data.

**data**                    Data frame with the omic identifiers as row names and just one column named Counts containing numeric values used as initial sample for the simulation.

**associationList**        Only for regulatory omics, a data frame with 2 columns, the first called containing the regulator ID and the second called Gene with the gene identifier.

**Value**

Initialized simulation object with the given data.

**Examples**

```
# Take a subset of the included dataset for illustration
# purposes. We could also load it from a csv file or RData,
# as long as we transform it to have 1 column named "Counts"
# and the identifiers as row names.

data(sampleData)

custom_rnaseq <- head(sampleData$SimRNaseq$data, 100)

# In this case, 'custom_rnaseq' is a data frame with
# the structure:
head(custom_rnaseq)
##           Counts
## ENSMUSG0000000001 6572
## ENSMUSG0000000003    0
## ENSMUSG0000000028 4644
## ENSMUSG0000000031    8
## ENSMUSG0000000037    0
## ENSMUSG0000000049    0

# The helper 'omicData' returns an object with our custom data.
rnaseq_customdata <- omicData("RNA-seq", data = custom_rnaseq)
```

---

omicResults

*Retrieves the simulated data.*

---

**Description**

Retrieves the simulated data.

**Usage**

```
omicResults(simulation, omics = NULL, format = "data.frame")
```

**Arguments**

simulation	A MOSimulation object.
omics	List of the omics to retrieve the simulated data.
format	Type of object to use for returning the results

**Value**

A list containing an element for every omic specific, with the simulation data in the format indicated, or a numeric matrix with simulated data if the omic name is directly provided.

**Examples**

```
omic_list <- c("RNA-seq")
rnaseq_simulation <- mosim(omics = omic_list)
#' # This will be a data frame with RNA-seq counts
rnaseq_simulated <- omicResults(rnaseq_simulation, "RNA-seq")
```

```
#           Group1.Time0.Rep1 Group1.Time0.Rep2 Group1.Time0.Rep3 ...
# ENSMUSG00000073155           4539           5374           5808 ...
# ENSMUSG00000026251              0              0              0 ...
# ENSMUSG00000040472           2742           2714           2912 ...
# ENSMUSG00000021598           5256           4640           5130 ...
# ENSMUSG00000032348            421            348            492 ...
# ENSMUSG00000097226             16             14             9 ...
# ENSMUSG00000027857              0              0              0 ...
# ENSMUSG00000032081              1              0              0 ...
# ENSMUSG00000097164            794            822            965 ...
# ENSMUSG00000097871              0              0              0 ...
```

---

omicSettings

*Retrieves the settings used in a simulation*

---

**Description**

Retrieves the settings used in a simulation

**Usage**

```
omicSettings(simulation, omics = NULL, association = FALSE,
             reverse = FALSE, only.linked = FALSE, prefix = FALSE,
             include.lagged = TRUE)
```

**Arguments**

simulation	A MOSimulation object.
omics	List of omics to retrieve the settings.
association	A boolean indicating if the association must also be returned for the regulators.
reverse	A boolean, swap the column order in the association list in case we want to use the output directly and the program requires a different ordering.

<code>only.linked</code>	Return only the interactions that have an effect.
<code>prefix</code>	Logical indicating if the name of the omic should prefix the name of the regulator.
<code>include.lagged</code>	Logical indicating if interactions with transitory profile and different minimum/maximum time point between gene and regulator should be included or not.

### Value

A list containing a data frame with the settings used to simulate each of the indicated omics. If association is TRUE, it will be a list with 3 keys: 'associations', 'settings' and 'regulators', with the first two keys being a list containing the information for the selected omics and the last one a global data frame giving the merged information.

### Examples

```
omic_list <- c("RNA-seq", "miRNA-seq")
multi_simulation <- mosim(omics = omic_list)

# This will be a data frame with RNA-seq settings (DE flag, profiles)
rnaseq_settings <- omicSettings(multi_simulation, "RNA-seq")

# This will be a list containing all the simulated omics (RNA-seq
# and DNase-seq in this case)
all_settings <- omicSettings(multi_simulation)
```

---

omicSim	<i>Set the simulation settings for an omic.</i>
---------	---

---

### Description

Set the simulation settings for an omic.

### Usage

```
omicSim(omic, depth = NULL, totalFeatures = NULL,
        regulatorEffect = NULL)
```

### Arguments

<code>omic</code>	Name of the omic to set the settings.
<code>depth</code>	Sequencing depth in millions of counts. If not provided will take the global parameter passed to mosim function.
<code>totalFeatures</code>	Limit the number of features to simulate. By default include all present in the dataset.
<code>regulatorEffect</code>	only for regulatory omics. Associative list containing the percentage of effects over the total number of regulator, including repressor, association and no effect (NE).

**Value**

A list with the appropriate structure to be given as options in mosim function.

**Examples**

```
omic_list <- c("RNA-seq")

rnaseq_options <- omicSim("RNA-seq", totalFeatures = 2500)

# The return value is an associative list compatible with
# 'omicsOptions'
rnaseq_simulation <- mosim(omics = omic_list,
                          omicsOptions = rnaseq_options)
```

---

plotProfile

*Generate a plot of a feature's profile for one or two omics.*

---

**Description**

Generate a plot of a feature's profile for one or two omics.

**Usage**

```
plotProfile(simulation, omics, featureIDS, drawReps = FALSE,
            groups = NULL)
```

**Arguments**

simulation	A MOSimulation object
omics	Character vector of the omics to simulate.
featureIDS	List containing the feature to show per omic. Must have the omics as the list names and the features as values.
drawReps	Logical to enable/disable the representation of the replicates inside the plot.
groups	Character vector indicating the groups to plot in the form "GroupX" (i.e. Group1)

**Value**

A ggplot2 object.

**Examples**

```
omic_list <- c("RNA-seq", "miRNA-seq")
rnaseq_simulation <- mosim(omics = omic_list)

plotProfile(rnaseq_simulation,
            omics = c("RNA-seq", "miRNA-seq"),
            featureIDS = list("RNA-seq"="ENSMUSG0000007682", "miRNA-seq"="mmu-miR-320-3p")
            )
```



---

`sampleData`*Default data*

---

**Description**

Dataset with base counts and id-gene tables.

**Usage**

```
sampleData
```

**Format**

An object of class `list` of length 6.

**Details**

List with 6 elements:

**SimRNAseq data** Dataframe with base counts with gene id as rownames.

**geneLength** Length of every gene.

**SimChIPseq data** Dataframe with base counts with regions as rownames.

**idToGene** Dataframe with region as "ID" column and gene name on "Gene" column.

**SimDNaseseq data** Dataframe with base counts with regions as rownames.

**idToGene** Dataframe with region as "ID" column and gene name on "Gene" column.

**SimMiRNAseq data** Dataframe with base counts with miRNA id as rownames.

**idToGene** Dataframe with miRNA as "ID" column and gene name on "Gene" column.

**SimMethylseq idToGene** Dataframe with region as "ID" column and gene name on "Gene" column.

**CpGisland** Dataframe of CpG to be used as initialization data, located on "Region" column

# Index

## \*Topic **datasets**

sampleData, 9

experimentalDesign, 2

is.declared, 3

mosim, 3, 4

MOSim-package, 2

omicData, 3, 4

omicResults, 5

omicSettings, 6

omicSim, 3, 7

plotProfile, 8

sampleData, 9