# SMGR: Setup and Introduction

*SMGR (Joint statistical method for integrative analysis of single-cell multi-omics data)*

Vignette built on April 20, 2022.

**SMGR workflow**

This tutorial goes through the steps in the SMGR workflow:

- Identify conistent signals of genes and peaks based on scRNA-seq & scATAC-seq data.
- Reveal the latent representation of co-expressed genes and peaks, i.e. co-regulatory program
- Identify co-expressed TF and target genes, based on single-cell multi-omics analysis

The setup and running steps are shown as follows in this tutorial:

**Requirements**

- Input: expression matrix The input to SMGR is a list of scRNA-seq expression matrix and scATAC-seq matrix:

*For scRNA-seq matrix:* Each column corresponds to a sample (cell) and each row corresponds to a gene. Expression units: The preferred expression values are gene-summarized raw counts. Other measurements (such as FPKM) are also accepted.

*For scATAC-seq matrix:* Each column corresponds to a sample (cell) and each row corresponds to a peak.

# Installation

The R implementation of SMGR is based on R packages as below.

Therefore, you will need to install these packages, and some extra dependencies, to run SMGR:

```
if (!requireNamespace("BiocManager", quietly = TRUE)) install.packages("BiocManager")
BiocManager::version()
```

```
## [1] '3.13'
```

```
# If your bioconductor version is previous to 4.0, see the section bellow
# Required
packages = c('glmnet','MASS','purrr','mpath','zic','pscl','parallel')
# BiocManager::install(packages)
```

Now you are ready to install SMGR:

```
# devtools::install_github("QSong-github/SMGR")
library('SMGR')
```

*Note: please invoke these packages to run the following codes:*

```
pkg = c('glmnet','MASS','purrr','mpath','zic','pscl','parallel')
sapply(pkg, require, character.only = TRUE)
```

```
##   glmnet     MASS    purrr    mpath      zic     pscl parallel
##     TRUE     TRUE     TRUE     TRUE     TRUE     TRUE     TRUE
```

# Run SMGR with example data

Example data is deposited in the data folder.

```
# @param rna.cts simulated scRNA-seq data
rna.cts <- readRDS('simuation_scRNA-seq.RDS')

# @param atac.cts simulated scATAC-seq data
atac.cts <- readRDS(file='simuation_scATAC-seq.RDS')

input_data <- list(rna.cts, atac.cts)
```

### SMGR process

Input data is a list of scRNA-seq and scATAC-seq data

```
input_data <- list(as.matrix(rna.cts),as.matrix(atac.cts))

#result1 <- smgr_main(sm.data = input_data, K=nrow(input_data[[1]]))
```

*if output is set to 'all', you will obtain the coefficients and BIC values*