

# Running multi-omics Integrating methods with CrIMMix

*Morgane Pierre-Jean*

*2019-09-12*

This vignette aims to show how to use the package named **CrIMMix**. This package groups several R packages that perform integrating analysis of multi-omics datasets. In this package, we can also simulate data in order to perform a comparative analysis with existent packages.

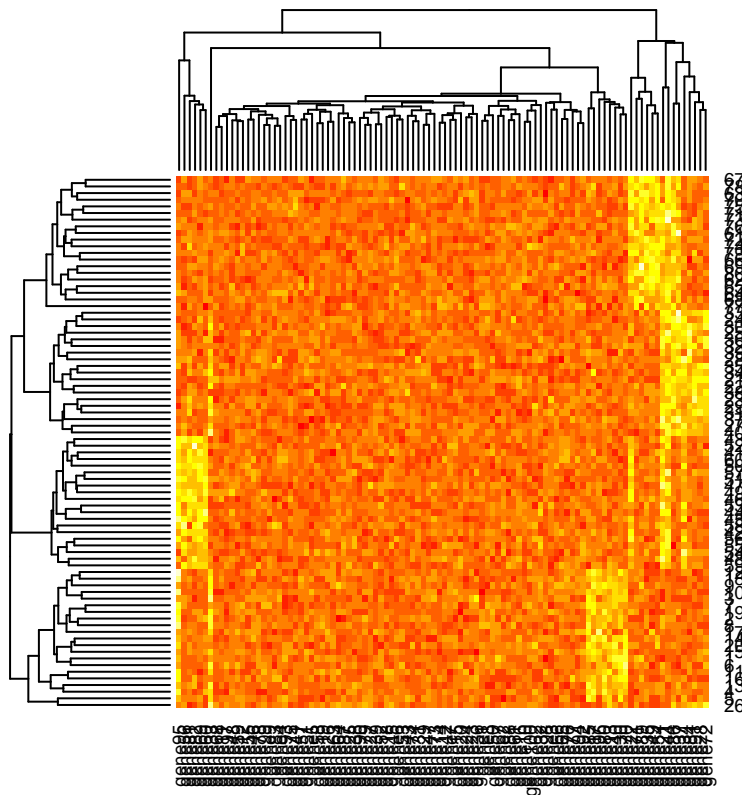
## Simulate Datasets

It is possible to tune the number of clusters, the number of samples per clusters, the length of signals. Then, the users can define the distribution of the data between the normal, the binary, or the uniform distribution.

### Normal distribution

The simulation under normal distribution could represent for instance gene expression data or raw copy number signal. For each cluster, the user can choose the mean and the variance of the distribution that varies from a cluster to another. The user can also choose the proportion of variables affected by this modified distribution. To finish, it is possible to modify the background noise.

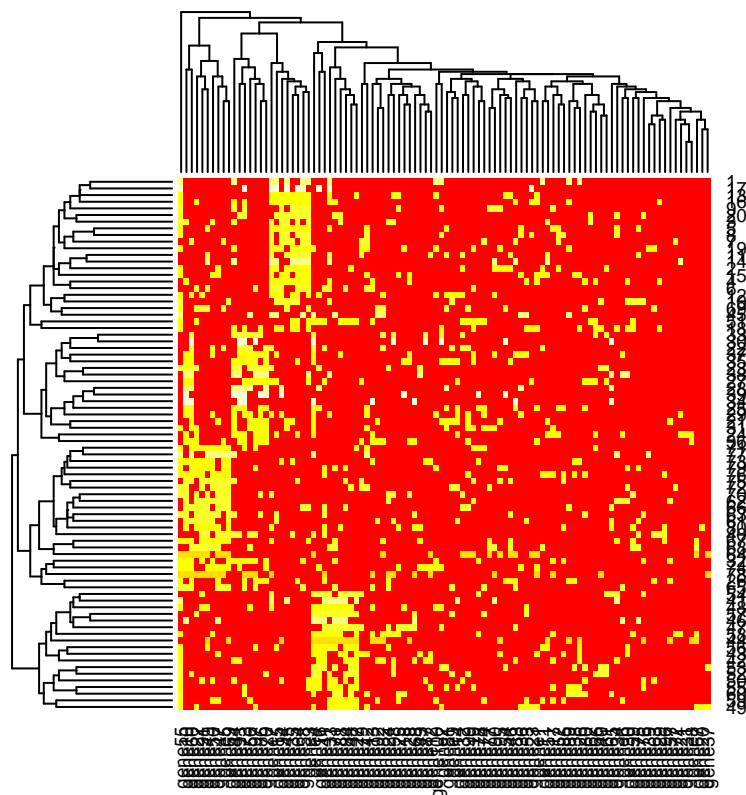
```
set.seed(34)
c_1 <- simulateY(nclust=4, n_byClust=20, J=100, flavor="normal", params=list(c(mean=3,sd=1)), prop=0.1,
heatmap(c_1$data)
```



## Binary distribution

The simulation under binary distribution could represent for instance mutations in tumor samples. For each cluster, the user can choose the mutation rate that varies from a cluster to another. The user can also choose the proportion of variables affected by this modified distribution. To finish, it is possible to modify the background noise (added random mutation at a proportion equal to the noise).

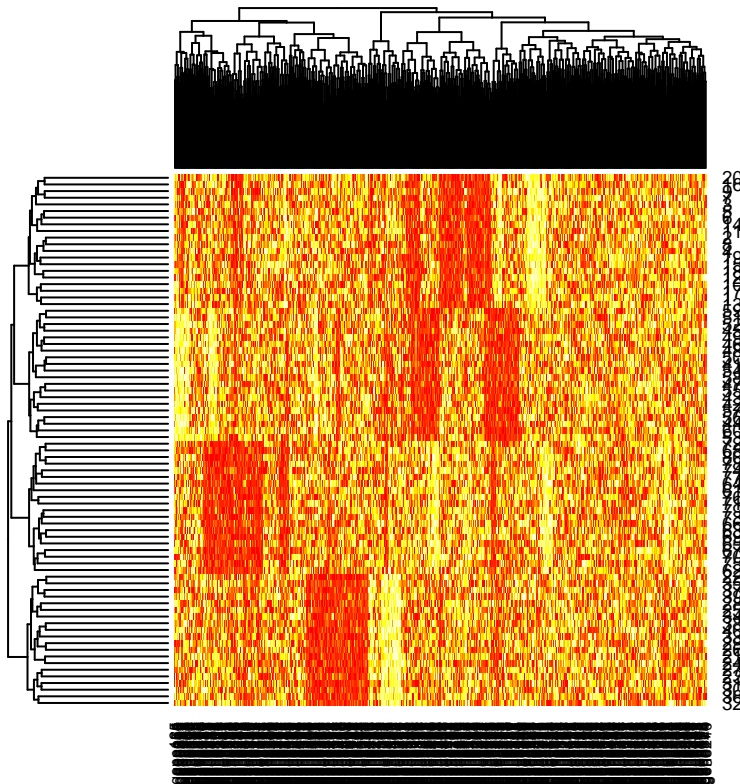
```
set.seed(34)
c_2 <- simulateY(nclust=4, n_byClust=20, J=100, flavor="binary", params=list(c(p=0.7)), prop=0.1, noise=0.1)
heatmap(c_2$data)
```



## Beta distribution

The simulation under uniform distribution could represent for instance methylation data. The beta distribution simulates data between 0 and 1. The user can also choose the proportion of variables that are hypermethylated or hypomethylated for the same subgroup. To finish, it is possible to modify the background noise.

```
set.seed(34)
params_beta <- list(c(mean1=-2, mean2=2, sd1=0.5, sd2=0.5))
c_3 <- simulateY(nclust=4, n_byClust=20, J=500, flavor="beta", params=params_beta, prop=0.2, noise=1)
heatmap(c_3$data)
```



## Run Integrative methods with CrIMMix

We create a wrapper called `IntMultiOmics` to run simply several methods dealing with multi-omics data. The user can call the method with the argument named `method` in function `IntMultiOmics`. It is possible to pass various arguments that are specific to each method. If nothing is specified, the default parameters of each method are used. Below, we give examples to run functions: SNF, RGCCA, and iCluster. For each method, we compute the ARI (Adjusted Rand Index).

```
data <- list(c_1$data, c_2$data, c_3$data)
```

### Run Mocluster

An example of how to run Mocluster Meng et al. (2015) methods is given below:

```
res_Mocluster <- IntMultiOmics(data, K=4, method="Mocluster", k=c(0.2, 0.2, 0.4))
```

### Run RGCCA

Then, we run RGCCA Tenenhaus and Tenenhaus (2011).

```
res_RGCCA <- IntMultiOmics(data, K=4, method="RGCCA")
```

### Run iCluster

To finish we try iCluster Mo et al. (2013).

```
res_icluster <- IntMultiOmics(data, K=4, method="iCluster", type=c("gaussian", "binomial", "gaussian"),
```

## Performance Evaluation

### Clustering evaluation with ARI

```
adjustedRIComputing(res_Mocluster,c_1$true.clusters)
#> [1] 1
```

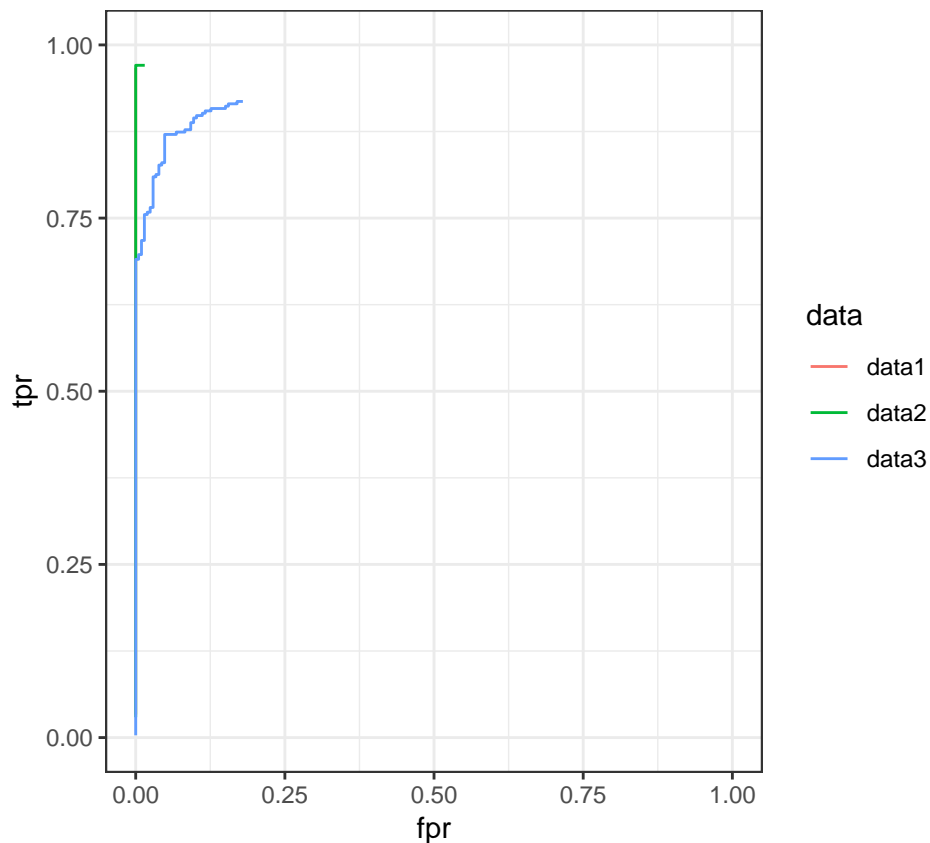
```
adjustedRIComputing(res_RGCCA,c_1$true.clusters)
#> [1] 0.721469
```

```
adjustedRIComputing(res_icluster,c_1$true.clusters)
#> [1] 0.9102489
```

### ROC evaluation

```
trueDat1 <- c_1$positive %>% unlist %>% unique
trueDat2 <- c_2$positive %>% unlist %>% unique
trueDat3 <- c_3$positive %>% unlist %>% unique
truth <- list(trueDat1,trueDat2, trueDat3)
```

```
roc_moclust <- roc_eval(truth= truth, fit = res_Mocluster$fit, method = "Mocluster")
plot_roc_eval(roc_moclust)
```



## References

- Meng, Chen, Dominic Helm, Martin Frejno, and Bernhard Kuster. 2015. “Mocluster: Identifying Joint Patterns Across Multiple Omics Data Sets.” *Journal of Proteome Research* 15 (3). ACS Publications: 755–65.
- Mo, Qianxing, Sijian Wang, Venkatraman E Seshan, Adam B Olshen, Nikolaus Schultz, Chris Sander, R Scott Powers, Marc Ladanyi, and Ronglai Shen. 2013. “Pattern Discovery and Cancer Gene Identification in Integrated Cancer Genomic Data.” *Proceedings of the National Academy of Sciences* 110 (11). National Acad Sciences: 4245–50.
- Tenenhaus, Arthur, and Michel Tenenhaus. 2011. “Regularized Generalized Canonical Correlation Analysis.” *Psychometrika* 76 (2). Springer: 257–84.