

# Introduction to SCRABBLE

*true*

*Updated on 2019-03-06*

Package version: SCRABBLE

## Contents

---

<b>1 Simulation Data Study</b>	<b>1</b>
1.1 Data Generation . . . . .	1
1.2 Run SCRABBLE . . . . .	2
<b>2 SessionInfo</b>	<b>3</b>
<code>library(SCRABBLE)</code>	

## 1 Simulation Data Study

---

### 1.1 Data Generation

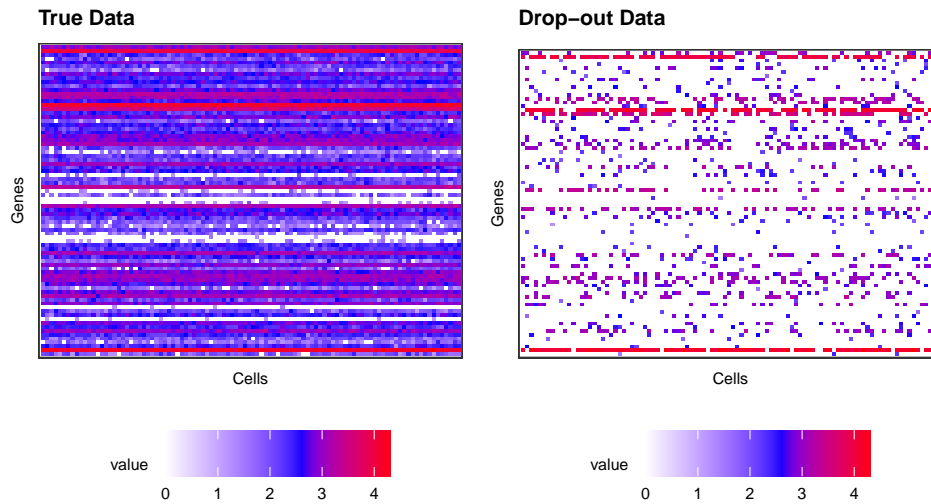
We simulated scRNA-Seq data consisting of three cell types using the Splat method in the Bioconductor package Splatter 8. Dataset consists of 800 genes and 1000 cells. The details of the parameters used in the simulation data generation are seen in Supplementary Table 2. Dropout midpoints (parameter `dropout_mid` in Splatter) are used to control the dropout rate in the simulated data. Splatter generates the data matrices for the true data and its corresponding dropout data using a given dropout midpoint. The corresponding bulk RNA-Seq data are the mean values of genes in the true scRNA-Seq data. The dropout RNA-Seq and bulk RNA-Seq data matrices are the inputs of the imputation methods. To determine the performance stability of the methods, we generated 100 datasets for each dropout midpoints.

#### 1.1.1 Load the data

```
data_sc <- demo_data[[1]]
data_bulk <- demo_data[[2]]
data_true <- demo_data[[3]]
```

#### 1.1.2 Plot the data

```
p1 <- list()
p1[[1]] <- plot_data(log10(data_true + 1), "True Data")
#> [1] 80 100
p1[[2]] <- plot_data(log10(data_sc + 1), "Drop-out Data")
#> [1] 80 100
main <- gridExtra::grid.arrange(grobs = p1, ncol = 2, top = "")
```



## 1.2 Run SCRABBLE

This package imputes drop-out data by optimizing an objective function that consists of three terms. The first term ensures that imputed values for genes with nonzero expression remain as close to their original values as possible, thus minimizing unwanted bias towards expressed genes. The second term ensures the rank of the imputed data matrix to be as small as possible. The rationale is that we only expect a limited number of distinct cell types in the samples. The third term operates on the bulk RNA-Seq data. It ensures consistency between the average gene expression of the aggregated imputed data and the average gene expression of the bulk RNA-Seq data. We developed 58 a convex optimization algorithm to minimize the objective function.

### 1.2.1 Set up the parameter used in SCRABBLE

```
parameter <- c(1,1e-6,1e-4)
```

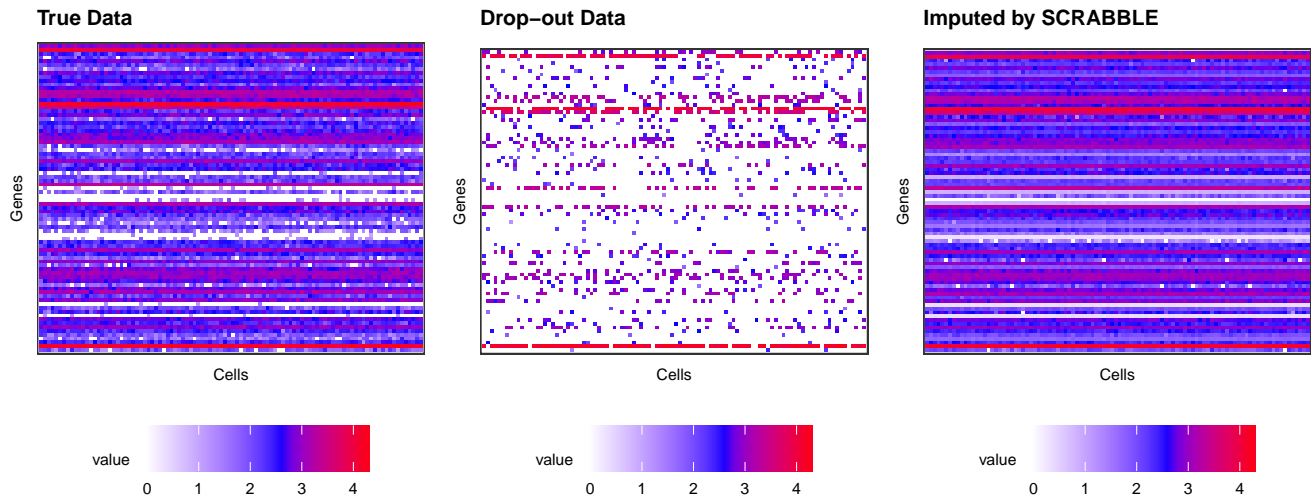
### 1.2.2 Run SCRABBLE

```
result <- scrabble(demo_data, parameter = parameter)
#> [1] "SCRABBLE begins the imputation of the data with 80 genes and 100 cells"
#> [1] "Imputation initialization is finished"
#> [1] "... .."
#> [1] "Imputation is finished"
```

### 1.2.3 Plot the data

```
p1 <- list()
p1[[1]] <- plot_data(log10(demo_data[[3]] + 1), "True Data")
#> [1] 80 100
p1[[2]] <- plot_data(log10(demo_data[[1]] + 1), "Drop-out Data")
#> [1] 80 100
p1[[3]] <- plot_data(log10(result + 1), "Imputed by SCRABBLE")
```

```
#> [1] 80 100
main <- gridExtra::grid.arrange(grobs = pl, ncol = 3, top = "")
```



## 2 SessionInfo

```
sessionInfo()
#> R version 3.4.4 (2018-03-15)
#> Platform: x86_64-apple-darwin15.6.0 (64-bit)
#> Running under: macOS 10.14.2
#>
#> Matrix products: default
#> BLAS: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRblas.0.dylib
#> LAPACK: /Library/Frameworks/R.framework/Versions/3.4/Resources/lib/libRlapack.dylib
#>
#> locale:
#> [1] C/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
#>
#> attached base packages:
#> [1] stats    graphics grDevices utils      datasets methods  base
#>
#> other attached packages:
#> [1] SCRABBLE_0.0.1 BiocStyle_2.4.1
#>
#> loaded via a namespace (and not attached):
#> [1] Rcpp_0.12.18      knitr_1.20        magrittr_1.5
#> [4] munsell_0.4.3     colorspace_1.3-2 lattice_0.20-35
#> [7] rlang_0.2.0       stringr_1.2.0     plyr_1.8.4
#> [10] tools_3.4.4       grid_3.4.4        gtable_0.2.0
#> [13] RSpecra_0.13-1    htmltools_0.3.6   rARPACK_0.11-0
#> [16] yaml_2.2.0        lazyeval_0.2.1    digest_0.6.15
#> [19] tibble_1.4.2      Matrix_1.2-12     gridExtra_2.3
#> [22] reshape2_1.4.3    RColorBrewer_1.1-2 ggplot2_2.2.1
#> [25] evaluate_0.10.1   rmarkdown_1.11    labeling_0.3
```

```
#> [28] stringi_1.1.6   pracma_2.1.8   pillar_1.2.1  
#> [31] compiler_3.4.4  scales_0.5.0
```