MRFscRNAseq Vignette

Hongyu Li

2021-04-22

Contents

Installing MRFscRNAseq from GitHub

```
if (!requireNamespace("devtools", quietly = TRUE))
    install.packages("devtools")
library(devtools)
install_github("eddiehli/MRFscRNAseq", quiet = TRUE, force = T)
library(MRFscRNAseq)
```

Main Function

The main function is

```
?get_DE_MRF()
```

It requires three input values: (1) data: a summary statistics matrix where rows are genes and columns are cell types; (2) g_g: the gene to gene binary network matrix; and (3) c_c: the cell type dependency binary network matrix. The usage of this function will be illustrated on a simulated data set.

Example

The `data_example` is a list object that contains the expression array, the gene to gene binary network matrix, and the cell type dependency binary network matrix. The expression array profiles 200 genes across 18 cell types in 2 groups. Each cell type has 200 cells.

```
data = MRFscRNAseq::data_example
gene_gene = data$gene_gene
gene_gene[1:10, 1:10]
#>     G1 G2 G3 G4 G5 G6 G7 G8 G9 G10
#> G1   0  0  0  0  0  1  0  0  0   0
#> G2   0  0  0  0  0  1  1  0  1   0
#> G3   0  0  0  0  0  1  0  1  0   0
#> G4   0  0  0  0  0  0  0  0  0   1
#> G5   0  0  0  0  0  0  0  1  1   0
#> G6   1  1  1  0  0  0  1  0  0   0
#> G7   0  1  0  0  0  1  0  0  0   1
```

```
#> G8   0  0  1  0  1  0  0  0  0   0
#> G9   0  1  0  0  1  0  0  0  0   0
#> G10  0  0  0  1  0  0  1  0  0   0
cell_cell = data$cell_cell
cell_cell[1:10, 1:10]
#>     C1 C2 C3 C4 C5 C6 C7 C8 C9 C10
#> C1   0  1  0  0  0  0  0  1  1   0
#> C2   1  0  0  0  0  0  0  1  1   0
#> C3   0  0  0  1  1  1  1  0  0   1
#> C4   0  0  1  0  1  1  1  0  0   1
#> C5   0  0  1  1  0  1  1  0  0   1
#> C6   0  0  1  1  1  0  1  0  0   1
#> C7   0  0  1  1  1  1  0  0  0   1
#> C8   1  1  0  0  0  0  0  0  1   0
#> C9   1  1  0  0  0  0  0  1  0   0
#> C10  0  0  1  1  1  1  1  0  0   0
```

Then we use our proposed MRF models to obtain DE results, we first covert the raw expression values to z scores using two-sample t-tests,

```
zz = MRFscRNAseq::get_z_scores(data)
```

Then we ran the MRF model. Note that here we set `iterEM = 20`, `iterGibbsPost = 500`, `brPost = 200` for illustrative purposes only. The default values are `iterEM = 200`, `iterGibbsPost = 20000`, `brPost = 10000`.

```
MRF_Results = get_DE_MRF(zz, gene_gene, cell_cell,
                         iterEM = 1, iterGibbsPost = 100, brPost = 50)
#>
#>  Estimating model parameters, 100 % completed
#>  Posterior sampling, 1 % completed Posterior sampling, 2 % completed Posterior sampling, 3 % complet
```

The MRF parameters are

```
MRF_Results$paraMRF
#>      Gamma  Beta_Gene  Beta_Cell
#> 0.11922946 0.06413417 0.05624906
```

We use this information to simulate a data set. The number of cell types was set to be 20. We set those genes that are assumed to be DE in the first 10 cell types to be DE. The rest were set to be equally expressed. Then the counts were simulated from a Negative Binomial distribution with mean $\alpha = 2$, $\beta = 4$ and dispersion $\phi = 0.1$, and $\tau = 5$.

First, we get p-values for the simulated data set using `ttest`, and obtain sensitivity, specificity and fdr for the results using significance level at $\alpha = 0.01$,