

User Manual of SingCellNet

Xiaomeng Zhang, Haifen Chen, Jie Zheng
School of Computer Science and Engineering
Nanyang Technological University, Singapore
Email: zhengjie@ntu.edu.sg

Last updated: Sept. 10, 2016

Launch the GUI

Open the file SingCellNet.m in MATLAB and press F5 to run the program. Then, the GUI of SingCellNet will be shown as in Figure 1.

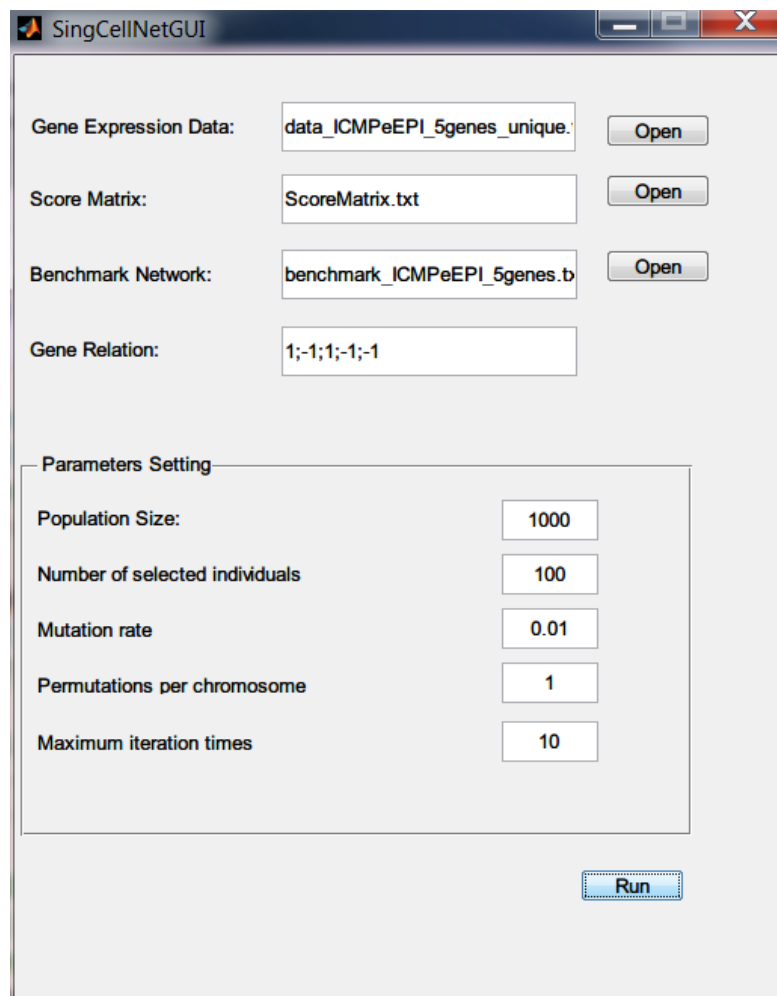


Figure 1. A screenshot of the GUI (graphical user interface) of SingCellNet.

Users can set the parameters of the program through the GUI. There are three mandatory parameters: 'Gene Expression Data', 'Score Matrix' and 'Gene Relation'. For other parameters, default values have been provided, but users are encouraged to try other values. After the parameters are set, click the "Run" button to execute the program.

Formats of files and parameters

"Gene expression data" file is a tab-separated txt file that contains two parts. There is an $n \times m$ matrix in the first part, where n is the number of cell types (i.e. nodes in the cell-lineage tree) and m is the number of genes in the dataset. The first column of the matrix contains the numbers of samples in corresponding cell types, while other entries are filled with the dummy value of 0. The second part of the file is an $S \times m$ matrix, where S is the number of samples in the experiment. Each entry of the matrix contains a binarized value of gene expression level. For example, if there are two cell types and three genes in the experiment, and there are two and three cells in the first and second cell types respectively, then the "Gene expression data" file may be as follows:

```
3  0  0
2  0  0
0  0  1
0  1  0
0  1  1
1  0  0
1  0  1.
```

In the above example, there are two cell types, where 3 cells are of type 1, and 2 cells are of type 2. The expression levels (in binary values) for the 5 cells are laid out in the following 5 rows.

"Score matrix" file is also a tab-separated txt file. It contains an $n \times n$ matrix, where n is the number of cell types. Each entry of the matrix contains the value of bonus or penalty for the transition of cell state from one cell type (row) to another cell type (column). For example, if we know differentiation has the directionality from cell type 1 to cell type 2, we can give such a transition a bonus of 3, while set a penalty of -2 for the transition in the opposite direction. In addition, the transition within the same cell type is given a neutral score of 1. As such, the score matrix is defined as follows.

```
1  3
-2 1
```

“Gene Relation” is an m -dimensional vector, where m is the number of genes and the entries are separated by semicolons. Each entry of the vector contains 1 or -1, which corresponds to the distribution of gene expression among cells. Here we assume that relations among genes (in terms of correlation of expression levels among cells) is “transitive”, e.g., if genes a and b are positively correlated, and genes b and c are positively correlated, then it is certain that a and c are also positively correlated. With such a simplification, then genes can be classified into 1 or -1, and the relation between any two genes can be figured out simply by looking up the vector. We are aware that this is an over-simplification, and will replace the vector with an $m \times m$ matrix of correlation among genes soon.

“Benchmark Network” file contains an $m \times m$ matrix \mathbf{B} , where m is the number of genes. If the i -th gene regulates the j -th gene, $\mathbf{B}[i][j] = 1$; otherwise, $\mathbf{B}[i][j] = 0$.