**---------------------------------------------------------------------------**

**ReadMe File**

**---------------------------------------------------------------------------**

**Prioritizing cancer-related microRNAs by integrating microRNA and mRNA datasets**

**Daeyong Jin, Hyunju Lee**

Data Mining & Computational Biology Lab

Gwangju Institute of Science & Technology

Gwangju, South Korea

http://combio.gist.ac.kr/

**---------------------------------------------------------------------------**

**I. List of input files**

**\* Please notice that all data in this page is simulation data. Our data just shows input format. We originally used expression data from TCGA (The Cancer Genome Atlas) and gene-miRNA prediction pairs from Pictar, targetScans and microCosm.**

**1. “./1\_compute\_feature\_values /data"**

**2. "./2\_integrate\_feature\_values /data"**

**II. List of Output files**

**1. “./1\_compute\_feature\_values /result"**

**2. "./2\_integrate\_feature\_values /result"**

**III. List of Source Codes**

**1. "./1\_compute\_feature\_values /src"**

**2. "./2\_integrate\_feature\_values /src"**

**---------------------------------------------------------------------------**

**Description**

**----------------------------------------------------------------------------------**

**I. Compute each feature value (F1, F2 and F3).**

**Source: "./1\_compute\_feature\_values /src"**

**Input : “./1\_compute\_feature\_values /data"**

**Output : “./1\_compute\_feature\_values /result”**

**1. “Compute\_F1.r”**

Calculate F1, the average miRNA expressions of all tumor samples.

**(input)**

(1) “simulated\_mirna\_exp.txt”

MiRNA expression data from tumor samples (expression values > 0)

**(output**)

(1) “F1\_ave\_mirna\_exp.txt”

Each row represents each F1 values for each miRNA (The average miRNA expressions of all tumor samples)

**2. “Compute\_F2.r”**

Calculate F2, the average of absolute Pearson correlation coefficients (PCCs) between miRNA and all gene expressions.

**(input)**

(1) “simulated\_mirna\_exp.txt”

MiRNA expression data from tumor samples (expression values > 0)

(2) “simulated\_gene\_exp.txt”

Gene expression data from tumor samples (expression values > 0)

**\* used paired samples with same order**

**(output)**

(1) “F2\_ave\_cor.txt”

Each row represents each F2 values for each miRNA (the average of absolute Pearson correlation coefficients (PCCs) between miRNA and all gene expressions.)

**3. “Compute\_F3.r”**

Calculate F3, the number of the predicted targets for each miRNA. First, we extract gene-miRNA pairs from gene-miRNA prediction database such as pictar, targetScans and microCosm. Then we get the union set of gene-miRNA prediction pairs. Then we calculate the number of the predicted targets for each miRNA.

**(input)**

(1) “gene\_list.txt”

a list of genes in gene expression data we used

(2) “mirna\_list.txt”

A list of miRNAs in miRNA expression data we used

(3) “simulated\_GMI.txt”

A list of predicted gene-miRNA pairs

**(output)**

(1) “F3\_n\_predicted\_target.txt”

Each row represents each F3 values for each miRNA (the number of the predicted targets for each miRNA.)

**II. Integrating features values based on order statistics**

**Source: "./2\_integrate\_feature\_values/src"**

**Input : “./ 2\_integrate\_feature\_values/data"**

**Output : “./ 2\_integrate\_feature\_values/result”**

**1. “integration\_using\_order\_statistics.r”**

Integrate three feature values using order statistics in below figure.



**(input)**

(1) “F1\_ave\_mirna\_exp.txt”

The average miRNA expressions values of all tumor samples

(2) “F2\_ave\_cor.txt”

The average of absolute Pearson Correlation Coefficients (PCCs) between miRNA and all gene expressions.

(3) “F3\_n\_predicted\_target.txt”

The number of the predicted targets for each miRNA.

**(output)**

(1) “integration\_result.txt”

Integration result of three features. Final ranking, miRNA, Q1, RQ1, Q2, RQ2 and RQ are outputted.