Pseudocode of the algorithm constructing the miRNA-miRNA functional synergistic network (MFSN)

Input: pre-processed miRNA-target interaction data; functional categories considered; $G_0$, protein interaction network; $G_1$, random protein interaction networks; $PG$, the significant threshold of functional enrichment; $O_{min}$, the threshold for the minimum number of genes; $D_1$, the threshold for the maximum value of the minimum distances of each target to others in the same subset; $D_2$, the threshold for the length of characteristic paths; $PC$, the significant threshold for characteristic path length.

Output: $MM$, the list of miRNA significantly functional synergistic pair; $FM$, the list of gene sets in corresponding co-regulating functional modules for miRNA pairs in $MM$.

1. Calculate the number $T$ of miRNAs in miRNA-target interaction data;
2. Calculate the number $N$ of target genes in miRNA-target interaction data;
3. Calculate the number $I$ of functional categories;
4. Calculate the number $R$ of random protein interaction networks;
5. Initialize $MM$ as NULL;
6. Initialize $FM$ as NULL;
7. For $s = 1$ to $(T-1)$, Do
8.   Detect miRNA $s$ regulating target set $A$ using miRNA-target interaction data;
9.   For $t = (s+1)$ to $T$, Do
10.  Detect miRNA $t$ regulating target set $B$ using miRNA-target interaction data;
11.  Detect their co-regulating target subset $A \cap B$;
12.  If $|A \cap B| > O_{min}$, then
13.     For $i = 1$ to $I$, Do
14.       Compute the number $K_i$ of genes at functional category $i$;
15.       Identify genes $(A \cap B)_i$ involved in $A \cap B$ and annotated at functional category $i$;
16.       If $|A \cap B_i| > O_{min}$, then
17.           Compute the probabilities $PG_i$ by hypergeometric distribution based on $N$, $K_i$, $|A \cap B|$, $|A \cap B_i|$;
18.           If $PG_i < PG$, then
19.             Calculate characteristic path length $CL$ of $(A \cap B)_i$ and the maximum of minimum distance MD of every target to others in $(A \cap B)_i$ in $G_0$:
20.             If $CL < D_2$ and $MD < D_1$, then
21.               Initialize $PC_{A \cap B_i} \rightarrow 0$;
22.               For $r = 1$ to $R$, Do
23.                 Calculate characteristic path length $CL_r$ of $(A \cap B)_i$ in $G_1$;
24.                 If $CL_r < CL$, then
25.                   $PG_{A \cap B_i} \leftarrow PG_{A \cap B_i} + 1/R$;
26.                 If $PC_{A \cap B_i} > PC$, then
27.                   Add $(s, t)$ to $MM$;
28.                   Add $(A \cap B_i)$ to $FM$;
29.       End For
30.   End For
31.   End For
32. End For
33. Output $MM$ and $FM$. 
Analysis of predicted miRNA targets from an integrated miRNA target data

Data source

In order to generate high efficient miRNA-target interactions, we integrate the miRNA targets predicted from 7 miRNA target data. MiRNA-target interactions are considered only when they occur in at least two data sources. The 7 data sources are TargetScan (version 5.1) (1), miRBase (version 5) (2), DIANA-microT (version 3.0, using the default loose score threshold) (3), PicTar (four-way) (4), miRanda (5) (September 2008 available at http://www.microrna.org/microrna/home.do), RNA22 (6), RNAhybrid (7) (downloaded from miRNAMap). Totally, there are 321,279 miRNA-target interactions between 784 miRNAs and 16,002 genes. Targets are represented by Entrez Gene IDs.

Identify miRNA pairs which significantly synergistically regulate functional modules and construct the MFSN

Theoretically speaking, 1,217,308,176 probabilities are computed between all pair combinations of the miRNAs (784*783/2) and all process categories considered (3966). Given $PG_i<0.005$, we detect 640,174 candidate functional modules and 8,726,253 different probabilities are computed totally. After two topological restrictions in protein interaction network with $PC<0.001$, 592 miRNAs are found to synergistically regulate 22,344 functional modules. On the basis of miRNA pairs regulating at least one functional module, we further construct the MFSN, containing 592 nodes and 6172 edges.

Properties of the MFSN

In the MFSN, all the 592 miRNAs are connected together (see Supp Fig1A), and the diameter is 2.6459 which is similar to those of random graphs generated by duplication model (2.9923+/-0.0951). In addition, the average clustering coefficient is also greater than it of random networks (0.1905 vs 0.0511+/-0.0137). So the network is a small-world network. Due to the small number of miRNAs which cooperate with several miRNAs, the distribution of degrees is not well following power-law (see Supp Fig1B). But, discarding these miRNAs whose degree is smaller than 4, the degree distribution reveals a power-law with a slope of -0.7129 and $R^2\approx 0.9319$. We
also observe with an increase of \( k \)-value, there is a sharp decrease in the number of cliques and 80.74% miRNAs occur in at least one clique (see Supp Fig1C). This result also reveals miRNAs neither as individual ones, nor big modules, but as small clusters to finish specific regulations together.

**Supp Fig1.** Figure of the MFSN and its structural features.

---

**Topological features of disease miRNAs**

From miR2Disease database, we totally obtain 305 miRNAs involved in 112 diseases as the “all disease data”, and the “high confidence disease data” containing 208 miRNAs correlated to 98 diseases. Next, we discuss whether they are close in the MFSN using the measure of characteristic path length among miRNAs for the same disease. Comparing to the two classes of control-miRNA sets, we find that the characteristic path length of the same disease miRNAs is significantly lower.
compared with permutation test without discern disease miRNAs \((P\text{ value}<0.00001)\), indicting that miRNAs for the same disease are closer to each other in the MFSN (see Supp Fig2A). This tendency also exists in “all disease data” (see Supp Fig2B). So we determine that miRNAs for the same disease tend to have direct or indirect but not distant functional synergism.

**Supp Fig2. **The mean characteristic path length among miRNAs for the same disease

In addition, we divide miRNAs into two groups: disease miRNAs and non-disease miRNAs. Then we calculate the significance of difference between the two groups and find disease miRNAs have statistically significantly higher degrees in the MFSN than non-disease miRNAs, where the median degree of disease miRNAs is 27 and non-disease miRNAs is 9 (see Supp Table1). We gain the same result in “all disease data”. In addition, we find an interesting result that the clustering coefficient of disease miRNAs is significantly larger than non-disease miRNAs (see Supp Table1). These results are accordant with those generated by TargetScan.

**Supp Table1** The significant difference of degrees and clustering coefficients in the MFSN contructed by miRBase between disease miRNAs and non-disease miRNAs using two kinds of disease data.

|                     | Degree | Clustering coefficient |
|---------------------|--------|------------------------|
|                     | D      | Non_D | P            | D | Non_D | P         |
| High confidence disease data | 27     | 9     | 1.21E-21     | 0.1795 | 0.1633 | 0.0073    |
| All disease data    | 24     | 8     | 3.35E-18     | 0.1778 | 0.1636 | 0.0412    |

\(D\) represents the median of degrees or clustering coefficients of disease miRNAs, \(\text{Non}_D\) is the median of degrees or clustering coefficients of non-disease miRNAs, \(P\) values are all calculated by wilcoxon rank sum test.

\(P\) values are all calculated by wilcoxon rank sum test.
Next, we explore the modular features of disease miRNAs without distinguishing classes of diseases. We also find that even though cluster sizes decrease with increasing $k$-value, the proportion of disease miRNAs identified in the protein communities increases, indicating the enrichment of disease miRNAs in the most tightly connected communities (see Supp Table2). We can get the same tendency using “all disease data”. These results show that disease miRNAs are tend to occur communities with high $k$ values. The tendency is found using TargetScan.

**Supp Table2 Number of communities and miRNAs at different $k$-values**

| $k$-value | comm_N | High confidence data | All disease data |
|-----------|---------|----------------------|------------------|
|           |         | Non_D_miRNA | D_miRNA | D_miRNA_ratio(%) | Non_D_miRNA | D_miRNA | D_miRNA_ratio(%) |
| 3         | 6       | 304          | 174     | 0.36402          | 241          | 237     | 0.4958          |
| 4         | 13      | 203          | 154     | 0.4314           | 155          | 202     | 0.5658           |
| 5         | 32      | 142          | 128     | 0.4741           | 103          | 167     | 0.6185           |
| 6         | 14      | 96           | 97      | 0.5026           | 68           | 125     | 0.6477           |
| 7         | 9       | 72           | 77      | 0.5168           | 52           | 97      | 0.6510           |
| 8         | 5       | 54           | 64      | 0.5424           | 38           | 80      | 0.6780           |
| 9         | 3       | 29           | 45      | 0.6081           | 22           | 52      | 0.7027           |
| 10        | 7       | 21           | 32      | 0.6038           | 15           | 38      | 0.7170           |
| 11        | 6       | 12           | 23      | 0.6571           | 10           | 25      | 0.7143           |

Comparing the disease miRNA set against the non-disease miRNA set reveals that disease miRNAs tend to reside at community interfaces compared with their non-disease counterparts in most kinds of communities, as shown in Supp Table3. Therefore disease miRNAs tend to work at the interface, which can be considered as the interface of multiple functions.

**Supp Table3 Multiple community membership distribution**

| $k$-value | High confidence disease data | All disease data |
|-----------|-------------------------------|------------------|
|           | Observed(%) | Expected(%) | Fold differ | Observed(%) | Expected(%) | Fold differ |
| 3         | 0.0227     | 0.0179     | 1.2662      | 0.0157     | 0.0199     | 0.7897     |
| 4         | 0.1600     | 0.0923     | 1.7354      | 0.1376     | 0.0927     | 1.4838     |
| 5         | 0.4426     | 0.4019     | 1.1013      | 0.4111     | 0.4111     | 1          |
| 6         | 0.2766     | 0.2808     | 0.9850      | 0.2899     | 0.2742     | 1.0571     |
| 7         | 0.2703     | 0.1964     | 1.3759      | 0.2692     | 0.1856     | 1.4509     |
| 8         | 0.0690     | 0.1798     | 0.3836      | 0.0732     | 0.1948     | 0.3756     |
| 9         | 0.4286     | 0.2075     | 2.0649      | 0.4138     | 0.1777     | 2.3276     |
| 10        | 0.6250     | 0.4054     | 1.5417      | 0.4783     | 0.4667     | 1.0248     |
| 11        | 0.6667     | 0.4783     | 1.3940      | 0.5625     | 0.5263     | 1.0688     |
Analysis of predicted miRNA targets from miRBase
Identify miRNA pairs which significantly synergistically regulate functional modules and construct the MFSN

Predicted miRNA target data is downloaded from miRBase database which uses miRanda algorithm. We consider human miRNAs, and convert Ensembl transcript IDs to Entrez Gene IDs. We obtain 401,862 regulations between 711 and 16,374 target genes, which is more than it obtained from TargetScan. Theoretically speaking, 982,865,070 probabilities are computed between all pair combinations of the miRNAs (711*710/2) and all process categories considered (3894). Given $PG_i<0.005$, we detect 348,033 candidate functional modules and 4,319,937 different probabilities are computed totally. After two topological restrictions in protein interaction network with $PC<0.001$, 692 miRNAs are found to synergistically regulate 10,415 functional modules. On the basis of miRNA pairs regulating at least one functional module, we further construct the MFSN, containing 692 nodes and 3872 edges.

Properties of the MFSN
In the MFSN, all the 692 miRNAs are connected together (see Supp Fig3A), and the diameter is 3.065 which is similar to those of random graphs generated by duplication model (3.101 +/- 0.101). In addition, the average clustering coefficient is also greater than it of random networks (0.0845 vs 0.0252 +/- 0.0082). So the network is a small-world network. Due to the small number of miRNAs which cooperate with several miRNAs, the distribution of degrees is not well following power-law (see Supp Fig3B). But, discarding these miRNAs whose degree is smaller than 4, the degree distribution reveals a power-law with a slope of -0.9995 and $R^2= 0.8462$. We also observe with an increase of k-value, there is a sharp decrease in the number of cliques and 76.16% miRNAs occur in at least one clique (see Supp Fig3C). This result also reveals miRNAs neither as individual ones, nor big modules, but as small clusters to finish specific regulations together.

Supp Fig3. Figure of the MFSN and its structural features.
Topological features of disease miRNAs

From miR2Disease database, we totally obtain 254 miRNAs involved in 109 diseases as the “all disease data”, and the “high confidence disease data” containing 174 miRNAs correlated to 96 diseases. Next, we discuss whether they are close in the MFSN using the measure of characteristic path length among miRNAs for the same disease. Comparing to the two classes of control-miRNA sets, we find that the characteristic path length of the same disease miRNAs is significantly lower (two \( P \) values <0.00001), indicting that miRNAs for the same disease are closer to each other in the MFSN (see Supp Fig4A). This tendency also exists in “all disease data” (see Supp Fig4B). So we determine that miRNAs for the same disease tend to have direct or indirect but not distant functional synergism.

**Supp Fig4.** The mean characteristic path length among miRNAs for the same disease is shorter than both two kinds of randomization tests.
In addition, we divide miRNAs into two groups: disease miRNAs and non-disease miRNAs. Then we calculate the significance of difference between the two groups and find disease miRNAs have statistically significantly higher degrees in the MFSN than non-disease miRNAs, where the median degree of disease miRNAs is 13 and non-disease miRNAs is 8 (see Supp Table4). We gain the same result in “all disease data”. In addition, we find an interesting result that the clustering coefficient of disease miRNAs is significantly larger than non-disease miRNAs (see Supp Table4). These results are accordant with those generated by TargetScan.

**Supp Table4** The significant difference of degrees and clustering coefficients in the MFSN constructed by miRBase between disease miRNAs and non-disease miRNAs using two kinds of disease data.

|                | Degree | Clustering coefficient |
|----------------|--------|------------------------|
|                | D      | Non_D | P            | D      | Non_D | P            |
| High confidence disease data | 13     | 8     | 2.63E-07     | 0.0765 | 0.0634 | 1.164E-05   |
| All disease data | 11     | 8     | 4.84E-05     | 0.0719 | 0.0632 | 0.0011      |

D represents the median of degrees or clustering coefficients of disease miRNAs, Non_D is the median of degrees or clustering coefficients of non-disease miRNAs, P values are all calculated by wilcoxon rank sum test.
P values are all calculated by wilcoxon rank sum test.

Next, we explore the modular features of disease miRNAs without distinguishing classes of diseases. We also find that even though cluster sizes decrease with increasing $k$-value, the proportion of disease miRNAs identified in the protein communities increases, indicating the enrichment of disease miRNAs in the most tightly connected communities (see Supp Table5). We can get the same tendency.
using “all disease data”. These results show that disease miRNAs are tend to occur communities with high k values. The tendency is found using TargetScan.

**Supp Table5 Number of communities and miRNAs at different k-values**

| k-value | comm_N | High confidence data | All disease data |
|---------|--------|----------------------|-----------------|
|         | Non_D_miRNA | D_miRNA | D_miRNA_ratio(%) | Non_D_miRNA | D_miRNA | D_miRNA_ratio(%) |
| 3       | 83     | 377       | 150             | 0.2846      | 315     | 212             | 0.4023 |
| 4       | 38     | 129       | 79              | 0.3798      | 109     | 99              | 0.4760 |
| 5       | 11     | 47        | 34              | 0.4198      | 41      | 40              | 0.4938 |
| 6       | 3      | 9         | 18              | 0.6667      | 8       | 19              | 0.7037 |
| 7       | 2      | 4         | 12              | 0.7500      | 4       | 12              | 0.7500 |

Because only two miRNAs beside interface of communities when k is 6 and there is no interface miRNA when k is 7, so these two kinds of communities are discarded below. Comparing the disease miRNA set against the non-disease miRNA set reveals that disease miRNAs reside at community interfaces to a much greater extent than their non-disease counterparts, as shown in Supp Table6. Therefore disease miRNAs tend to work at the interface, which can be considered as the interface of multiple functions.

**Supp Table6 Multiple community membership distribution**

| k-value | High confidence disease data | All disease data |
|---------|------------------------------|-----------------|
|         | Observed(%) | Expected(%) | Fold differ | Observed(%) | Expected(%) | Fold differ |
| 3       | 0.33333     | 0.2971      | 1.1220      | 0.3113      | 0.3048      | 1.0215      |
| 4       | 0.3165      | 0.2481      | 1.2757      | 0.2930      | 0.2569      | 1.1403      |
| 5       | 0.2647      | 0.1064      | 2.4882      | 0.2500      | 0.0976      | 2.5625      |
| 6       | 0           | 0.2222      | 0           | 0           | 0.2500      | 0           |
| 7       | 0           | 0           | 0           | 0           | 0           | 0           |

1. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. and Burge, C.B. (2003) Prediction of mammalian microRNA targets. *Cell*, **115**, 787-798.
2. Griffiths-Jones, S., Saini, H.K., van Dongen, S. and Enright, A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res*, **36**, D154-158.
3. Maragkakis, M., Reczko, M., Simossis, V.A., Alexiou, P., Papadopoulos, G.L., Dalamagas, T., Giannopoulos, G., Goumas, G., Koukis, E., Kourtis, K. et al. (2009) DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res*, **37**, W273-276.
4. Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. et al. (2005) Combinatorial microRNA target predictions. *Nat Genet*, **37**, 495-500.
5. John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C. and Marks, D.S. (2004) Human
MicroRNA targets. *PLoS Biol*, 2, e363.

6. Miranda, K.C., Huynh, T., Tay, Y., Ang, Y.S., Tam, W.L., Thomson, A.M., Lim, B. and Rigoutsos, I. (2006) A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes. *Cell*, 126, 1203-1217.

7. Rehmsmeier, M., Steffen, P., Hochsmann, M. and Giegerich, R. (2004) Fast and effective prediction of microRNA/target duplexes. *Rna*, 10, 1507-1517.