MiRNA–miRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features

Juan Xu, Chuan-Xing Li, Yong-Sheng Li, Jun-Ying Lv, Ye Ma, Ting-Ting Shao, Liang-De Xu, Ying-Ying Wang, Lei Du, Yun-Peng Zhang, Wei Jiang, Chun-Quan Li, Yun Xiao and Xia Li*

College of Bioinformatics Science and Technology, Harbin Medical University, Harbin 150081, China

ABSTRACT

Synergistic regulations among multiple microRNAs (miRNAs) are important to understand the mechanisms of complex post-transcriptional regulations in humans. Complex diseases are affected by several miRNAs rather than a single miRNA. So, it is a challenge to identify miRNA synergism and thereby further determine miRNA functions at a system-wide level and investigate disease miRNA features in the miRNA–miRNA synergistic network from a new view. Here, we constructed a miRNA–miRNA functional synergistic network (MFSN) via co-regulating functional modules that have three features: common targets of corresponding miRNA pairs, enriched in the same gene ontology category and close proximity in the protein interaction network. Predicted miRNA synergism is validated by significantly high co-expression of functional modules and significantly negative regulation to functional modules. We found that the MFSN exhibits a scale free, small world and modular architecture. Furthermore, the topological features of disease miRNAs in the MFSN are distinct from non-disease miRNAs. They have more synergism, indicating their higher complexity of functions and are the global central cores of the MFSN. In addition, miRNAs associated with the same disease are close to each other. The structure of the MFSN and the features of disease miRNAs are validated to be robust using different miRNA target data sets.

INTRODUCTION

MiRNAs are endogenous $\sim$22 nt small non-coding RNAs that repress gene expression by binding 3′-untranslated regions (UTRs) of mRNA target transcripts, causing translational repression or mRNA degradation. They guide many key biological processes and are involved in many diseases (1–3).

Some researchers claimed that the human genome might encode more than 1100 miRNAs (4,5) and it was recently demonstrated that they potentially regulate the majority of all human genes (6), which can therefore be used to improve our understanding of the mode of action of miRNAs and their functions (7). The limited miRNAs are thought to be able to control the larger set of genes through synergism, in which multiple miRNAs work synergistically to control individual genes. For example, lin-4 and let-7 are cooperative in *Drosophila* (8) and are the earliest miRNA pair to be experimentally verified. Krek *et al.* (9) also showed that miR-375, miR-124 and let-7b jointly regulate *Mtpn*, providing evidence for coordinate miRNA control in mammals. Interestingly, Wu *et al.* (10) recently found that 28 miRNAs can substantially inhibit p21Cip1/Waf1 expression, which ushers in a new era of miRNA research that focuses on networks more than on individual connections between miRNAs and strongly predicted targets. Studying the synergism of miRNAs is an important step for further determining miRNA functions at a system-wide level.

With the availability of large data sets derived from high-throughput experiments and computer algorithms, we could investigate the complex synergistic relationships between miRNAs. Our understanding of the synergistic regulation of miRNAs is increasing, and new methods are being developed to understand miRNA synergism

---

(11–15). Zhou *et al.* (13) used statistical measures to quantify the regulatory associations between the sets of predicted targets of miRNA pairs. A randomization-based test devised by Shalgi *et al.* (11) was used to identify miRNA pairs that exhibit significant co-occurrence in 3′-UTRs of the same target genes, while Boross *et al.* (12) constructed a miRNA co-regulation network by computing the correlations between the gene silencing scores of individual miRNAs. An *et al.* (14) used the signal-to-noise ratio to get high accurate regulating miRNAs for every gene and described a procedure to identify highly probable co-regulating miRNAs and the corresponding co-regulated gene groups, involving a sequence of statistical tests. DIANA-mirPath was developed to consider the combinatorial effect of co-expressed miRNAs in the modulation of a given pathway (15). All these studies have demonstrated the importance of miRNA synergism and also indicated that integrating predicted targets and functional information could identify synergistic miRNA pairs and simultaneously reveal their underlying functions.

In addition, the protein–protein interactions involve functional similarities between proteins. It has been found that interacting proteins are regulated by similar miRNA types (16), and clustered miRNAs also jointly regulate proteins that are close in the protein interaction network and the number of co-regulations between proteins is negatively correlated with their distance in the network (17). In addition, miRNAs can target gene batteries that are functionally related effector genes (18) and the target genes of miRNA clusters are enriched in communities in the protein interaction networks representing distinct cellular processes (17). Furthermore, integrating various types of data is one solution to decrease the false positives of high-throughput data. Another benefit is that it may enable us to consider many biological perspectives.

Based on the above observations, we developed a computational method to identify significantly functional synergistic miRNA pairs via functional modules that they jointly regulate by integrating predicted miRNA targets, their corresponding functional information and protein interaction data; these pairs are used to construct the miRNA–miRNA functional synergistic network (MFSN; our strategy is illustrated in Figure 1A). A functional module is a subset of genes that could independently implement a specific function as a whole. Here, it is defined as a subset of targets that satisfy three restrictions: co-regulated by miRNA pairs, enriched in the same gene ontology (GO) category and share close proximity in the protein interaction network. We also validated the predicted miRNA–miRNA synergism by co-expression of functional modules and their significant negative regulation to functional modules. Then, we analyzed the structural features of the MFSN using graph theoretical methods and found that it is a scale free, modular and small-world network. So, the network is a type of graph distinct from both regular and random networks, and can be used to further our understanding of miRNA functions

from a system-wide level. The structure of the MFSN is robust by using different miRNA target datasets.

In most conditions, miRNAs are synergistic in complex diseases and regulate genes with the same or similar functions. It is a challenge to understand the mechanism of diseases. For example, cardiac arrhymogenesis is linked to miR-1 and miR-133, both of which act through the regulation of essential ion channel proteins (19). Wu *et al.* also found that many of these 28 p21-regulating miRNAs are upregulated in cancers (10). These studies indicate that the synergism among miRNAs is important to understand the mechanisms of complex diseases. The increase in disease miRNA data also allows us to analyze their specific features at the system level. Here, we found that the topological features of disease miRNAs in the MFSN are significantly distinct from non-disease miRNAs. Disease miRNAs have more synergism than non-disease miRNAs, indicating that they have more complex functions. They are also the global central cores of the MFSN, indicating their greater centrality in the network. In addition, miRNAs associated with the same disease are located close to each other to allow regulation of the same or similar functions. We also validated that the features of disease miRNAs are accordant in different MFSNs using different microRNA target prediction algorithms.

MATERIALS AND METHODS

Data

Three types of human miRNA target data sets are analyzed. For our analyses, we used data from TargetScan and, as a control, from miRBase or integrated target data (detailed description about the two controls can be found in the Supplementary data). Targets predicted by TargetScan 5.1 (5) with a total context score of −0.30 or less are ignored, where the score can quantitatively measure the overall target efficacy (20,21). As controls, the highly efficient miRNA target data consisting of miRNA–target interactions that occur in at least two of seven considered sources and all the predicted human targets from miRBase (version v5) (22) are considered. The seven data sources are TargetScan 5.1 (5), miRBase (version v5) (22), DIANA-microT (version 3.0) (23), PicTar (four-way) (9), miRanda (24), RNA22 (25) and RNAhybrid (26). Protein interaction data are assembled from HPRD (HPRD_Release_8_070609) (27); here we only considered the maximum component of the whole protein interaction network, which contains 33 762 interactions between 8556 proteins. The GO file on Biological Processes is downloaded from the GO consortium (28), which is available at http://www.geneontology.org, as of November 2009. As in previous studies, process categories from GO are restricted to below the fourth level of the hierarchy to avoid analyzing very general non-descriptive terms (29,30). Finally, we used Entrez Gene IDs to represent corresponding genes or proteins.

The mRNA and miRNA microarray data of the NCI-60 are downloaded from the CellMiner database (31), which involves a panel of 60 human cancer cell
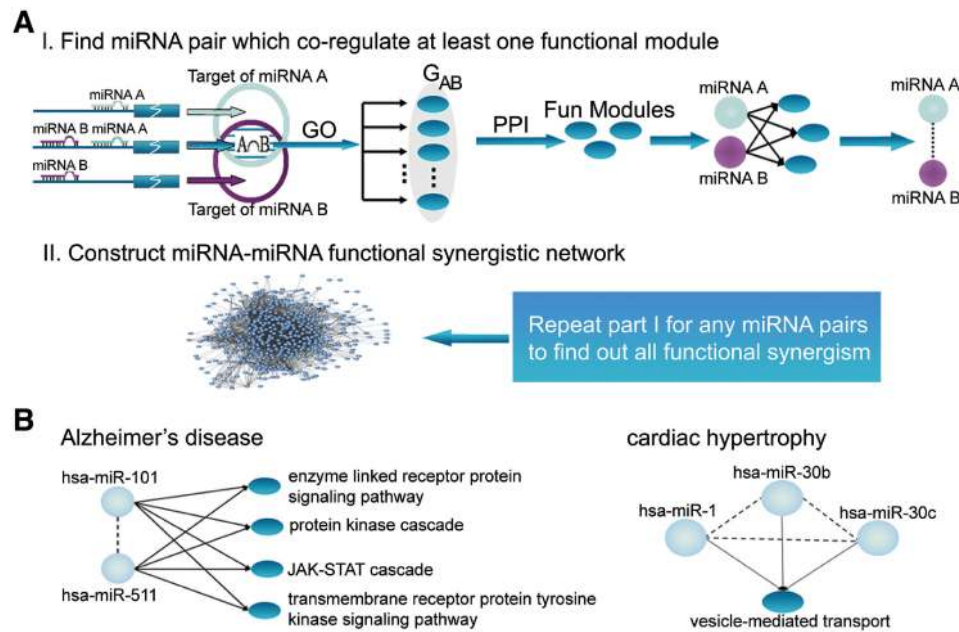
Figure 1. The workflow to construct the MFSN and two examples of synergism among miRNA with their co-regulating functional modules. (**A**) Workflow to construct the miRNA–miRNA functional synergistic network (MFSN) via co-regulating functional modules. The process involves two main steps. First, we identified an miRNA pair that synergistically regulates at least one functional module. Second, we repeated the first step for any miRNA pairs, and assembled all the significant miRNA pairs to construct the MFSN. (**B**) Two examples of miRNA pairs that synergistically regulate functional modules; these co-regulations are associated with diseases. Non-direct dashed line represents the miRNA synergistic action; direct line represents the miRNA regulation to the functional module.

lines from nine distinct tissues. The mRNA expression profiles are measured using the Affymetrix GeneChip HG-U133A platform, and we directly downloaded the normalized data set from the CellMiner database, where we selected the GCRMA algorithm and $\log_2$ transformation (32). GCRMA is a procedure of pre-processing oligonucleotide expression arrays using the robust multiarray average (RMA) with the help of probe sequence and with GC-content background correction (33). The miRNA data are collected using a miRNA OSU V3 chip and we downloaded data after $\log_2$ transformation. The miRBase database is then used to map miRNA probes to mature miRNAs (34). A total of 59 NCI samples are applied because we did not consider the NCI-H23 cell line that lacks mRNA data.

Information on disease miRNAs is obtained from the miR2Disease database (release 19 April 2009), which is a manually curated database that aims to provide a comprehensive record of miRNA deregulation involved in various human diseases (35). Based on the detection method of differential expression for miRNAs in diseases, we obtained two classes of disease data with different confidence levels. One is all the information that we could obtain from the database, defined as 'all disease data', whereas the other is the subset of data that is yielded by low-throughput methods such as northern blot and qRT–PCR approaches, and is termed the 'high confidence disease data'. The results for disease miRNAs discussed below are generated using the 'high confidence disease data', except where otherwise specified.

Methods

Identify miRNA pairs and construct miRNA–miRNA functional synergistic network

Overview the processes of identifying synergistic miRNA pairs. Figure 1A indicates the methodology used in this study. After we obtained miRNA target data sets from databases, the resulting data are preprocessed as described in the 'Data' section and the Supplementary data. First, for each miRNA pair, we identified their co-regulating targets as a target subset and identified candidate functional modules in the target subset by performing functional enrichment in each biological process category. When there is at least one candidate functional module, we used two topological features in the protein interaction network to filter functional modules in the candidate module set generated by functional enrichment. In this study, if a pair of miRNAs significantly co-regulates at least one functional module, we defined them as synergistic. Finally, the miRNA functional synergistic network is constructed by assembling all miRNA synergistic pairs, where nodes represent miRNAs and edges represent their functional synergism. In addition, the pseudocode of the algorithm is described in the Supplementary data.

For a given miRNA pair of miRNA A and B, we identified the target subset they co-regulate ($A \cap B$). The subset is required to have at least O_{min} genes. First, biological processes where the target subset is enriched are identified by hypergeometric distribution. The probability PG_i for $A \cap B$ in the GO term i is calculated

according to

$$PG_i = 1 - F(x|N,K_i,M) = 1 - \sum_{t=0}^{x} \frac{\binom{K_i}{t}\binom{N-K_i}{M-t}}{\binom{N}{M}}$$

$$(i = 1, 2, \ldots, I)$$

where N is the number of all targets (default background distribution), K_i is the total number of genes that are annotated in the GO term i and targeted by miRNAs, M is the size of $A \cap B$, x is the number of targets in $A \cap B$ that are also annotated to term i and I is the total number of GO terms we considered. At the given significance level, we not only obtained the set with enriched function terms but also captured the set G_{AB} with the subsets in $A \cap B$ that are annotated to each term in the previous set. Namely, G_{AB} is the set of candidate functional modules. Second, we further identified the functional modules in G_{AB}. We stipulated that a functional module must contain no fewer than O_{min} targets and simultaneously satisfy two topological restrictions in the protein interaction network: (i) the minimum distance from every target to others in the subset is no larger than the threshold $D1$; (ii) the characteristic path length is shorter than the threshold $D2$ and significantly shorter than random (the computing method of the characteristic path length is outlined in the next section). The significant P-value (PC) for the characteristic path length of modules in (ii) is calculated by using the edge-switching method and is defined as the fraction of characteristic path lengths of the same subset in random protein interaction networks that is shorter than that in the real network. Here, we generated 1000 random networks. As stringent controls, random networks are constructed by preserving the number of direct neighbors for each protein in the original protein interaction network using the edge switching method (36). Mfinder is used to generate random networks by selecting the options of the switching method to generate random networks and output all random networks (available at http://www.weizmann.ac.il/mcb/UriAlon/).

After performing the function enrichment and two topological restrictions in the protein interaction network, miRNA A and miRNA B are considered to be synergistic if they co-regulate at least one functional module. Here, the value of O_{min} is set to 3, and $D1$ is set to 2. To make the communication among nodes in a functional module quicker than under general conditions, we required that characteristic path lengths of modules are smaller than the diameter of the protein interaction network; this is because small diameter is a characteristic feature of small-world networks (37). Here, the threshold $D2$ is set to 4, smaller than the diameter of the protein interaction network (4.2327), making it a small-world biological network.

After assembling all significant miRNA pairs identified above, we generated the MFSN. A node represents a miRNA, and two nodes are connected if the corresponding miRNA pair has a synergistic action, otherwise no edge.

Topological measurements of the MFSN. In this study, we discussed several topological features at different levels. For the whole network, we identified its maximum component and calculated the diameter, which is defined as the average distance between any two nodes in the network. The distance between two nodes is the number of edges on a shortest path between them. We analyzed the degrees and clustering coefficients of nodes. For a given subset of nodes, we defined its characteristic path length as the average distance between the subset, and the same procedure is used to calculate the average distance of target subsets in the protein interaction network. To determine if the MFSN is a small-world network, we used the duplication model to construct random graphs; this is a well-known model having power-law degree distributions and providing small-world networks. We generated 10 000 instances and computed the average clustering coefficient and average diameter. The topological measurements and random networks are obtained using the 'RandomNetworks' plugin (beta version) of Cytoscape (38) and the matlab_bgl package (available from http://www.stanford.edu/~dgleich/programs/matlab_bgl/).

Next, we used the clique percolation clustering method (39) to identify miRNA functional synergistic modules, which are defined as cliques. Cliques are all of complete subgraphs that are not parts of larger complete subgraphs (39). The algorithm identifies maximal complete subgraphs (cliques) in the network and then identifies 'communities' by performing standard component analysis of the clique–clique overlap. Thus, the resulting communities are allowed to have a degree of overlap, which is particularly advantageous because such methods have been demonstrated to be more suitable for identifying central nodes in networks compared with non-overlapping clustering algorithms. This procedure is performed using CFinder (40), which is a fast program for locating and visualizing overlapping.

Randomization test. We evaluated the significance of co-expression of functional modules or negative regulations from miRNA to functional modules by randomly selecting genes as functional modules. For each functional module, we randomly selected the same number genes and calculated correlation values of the gene set or between the gene set and miRNAs that regulate the functional module; the procedure is repeated 10 000 times. Then, we calculated the average of correlations in each random condition. The significant P-value is the fraction of the average correlations in random conditions, which is greater than the value in the real condition.

To determine the statistical significance of the close proximity of miRNAs involved in the same disease, we calculated the characteristic path length between them in the MFSN. We then randomly selected the same number of miRNAs from the miRNA background set and computed the characteristic path length as described above. We repeated this procedure 100 000 times. Diseases that contain at least two miRNAs are all analyzed. The P-value is the fraction of average characteristic path lengths for all diseases in random networks, which is

shorter than that in the real MFSN. Here, we discussed two classes of miRNA background sets: all miRNAs included in the maximum component of the network, the intersection set comprising miRNAs of the first group that are also disease miRNAs.

RESULTS

Evaluation of miRNA pairs that synergistically regulate functional modules

We obtained 185 773 regulations between 676 miRNAs and 15 829 target genes from the TargetScan database using a high threshold of context score. Theoretically, 888 416 100 probabilities are computed between all pair combinations of the miRNAs (676*675/2) and all process categories considered (3894). Given the significant level of functional enrichment, $PG_i < 0.05$, we detected 472 573 regulations between miRNA pairs and candidate functional modules and a total of 1 071 920 different probabilities are computed. After two topological restrictions in the protein interaction network and the significance level of characteristic path length set to $PC < 0.001$, 13 687 functional modules are regulated by 2937 non-redundant synergistic patterns among 473 miRNAs, where a miRNA pair might regulate several distinct functional modules.

First, we performed co-expression of functional modules to assess the validity of our miRNA synergism predictions. Our working hypothesis is that the expression profiles of functional modules controlled by corresponding miRNA pairs are more likely to be correlated and behave more similarly than those of randomly selected gene sets. So, if we observed significant co-expression of targets in functional modules, we inferred that miRNA pairs co-regulate these functional modules. We used the average of correlations of functional modules as the measure of similarity. The correlation of a functional module is defined as the average correlation coefficients between each gene pair in the modules. In addition, the background correlation is the average correlation coefficient of any gene pair in the entire genome. We collected expression profiles of the NCI-60, derived from cancers of nine tissue origins, and then calculated the correlation of functional modules. The average of correlations is 0.3028, which is significantly higher than that expected under random conditions (the value at random is 0.192 39, $P < 0.0001$, see 'Methods' section) or for general background (0.1923). Thus, we concluded that functional modules are highly co-expressed.

As a second independent evaluation of our miRNA synergism predictions, we used negative regulations of miRNA synergistic pairs to functional modules. Our hypothesis is that if functional modules are regulated by miRNA pairs, the regulations should be stronger than those of randomly selected genes and of other common targets of the corresponding miRNA pairs. We calculated the average of correlations (R_{AvgUnit}) between miRNAs and functional modules using the miRNA and mRNA expression profiles of the NCI-60. The correlation between miRNA and its regulating functional module is defined as the average correlation coefficients between miRNAs and each gene in the module. As a result, the average of correlations is significantly greater than random ($R_{\mathrm{AvgUnit}} = -0.1225$, the value at random is $-0.102\,37$, $P < 0.0001$, see the 'Methods' section), indicating stronger negative regulation to functional modules. We also computed the correlations (R_{AvgOther}) between miRNAs and other common targets than functional modules and found that the correlations between miRNAs and functional modules are also significantly lower than the correlations between other common targets ($R_{\mathrm{AvgOther}} = -0.1165$, $P_{t-\mathrm{test}} < 10^{-32}$ computed by the t-test).

These results indicate that miRNA pairs identified by our method simultaneously regulate targets in functional modules and cause co-expression of these targets. For example, hsa-miR-101 and hsa-miR-511 synergistically regulate four functional modules, all of which are involved in signal transduction (left panel of Figure 1B). The average correlation coefficient between hsa-miR-101 (or hsa-miR-511) and four functional modules is -0.2453 (or -0.1577), indicating that these functional modules are under strong negative regulation of the two miRNAs. Meanwhile, the average co-expression of four functional modules is 0.3155. We further analyzed the functional concordance of hsa-miR-101 and hsa-miR-511 using the 'meet/min' score, which is the number of functional modules they regulate together divided by the smaller number of function modules of these two miRNAs. The score reaches 0.5357, suggesting high functional concordance. Previous studies have shown that these two miRNAs are both involved in Alzheimer's disease (41). Similarly, hsa-miR-1, hsa-miR-30b and hsa-miR-30c synergistically regulate vesicle-mediated transport (right panel of Figure 1B), and they are synergistic. These regulations are also strongly negative (data not shown). The functional concordance score between hsa-miR-1 and hsa-miR-30b (or hsa-miR-30c) is 0.75, and the score of hsa-miR-30b and hsa-miR-30c is much higher, which are in the same miRNA family. Down-regulations of the three miRNAs have also been found in cardiac hypertrophy (42), and the biological process of vesicle-mediated transport they regulate is associated with this disease.

These results indicate the feasibility of our methods in identifying miRNA pairs through their synergistically regulating functional modules. Therefore, we constructed the MFSN based on the results above.

Properties of the MFSN

On the basis of miRNA pairs regulating at least one functional module, we constructed the MFSN containing 473 nodes and 2937 edges (Figure 2A). The MFSN is an objective representation of all synergistic associations between miRNAs. The number of miRNAs in the network is four times more than previous studies. Next, we discussed the structure and organization of this network. From Figure 2B, we could see that a few miRNAs interact with a relatively large number of miRNA partners, whereas many miRNAs have few miRNA partners. The examination of the degree
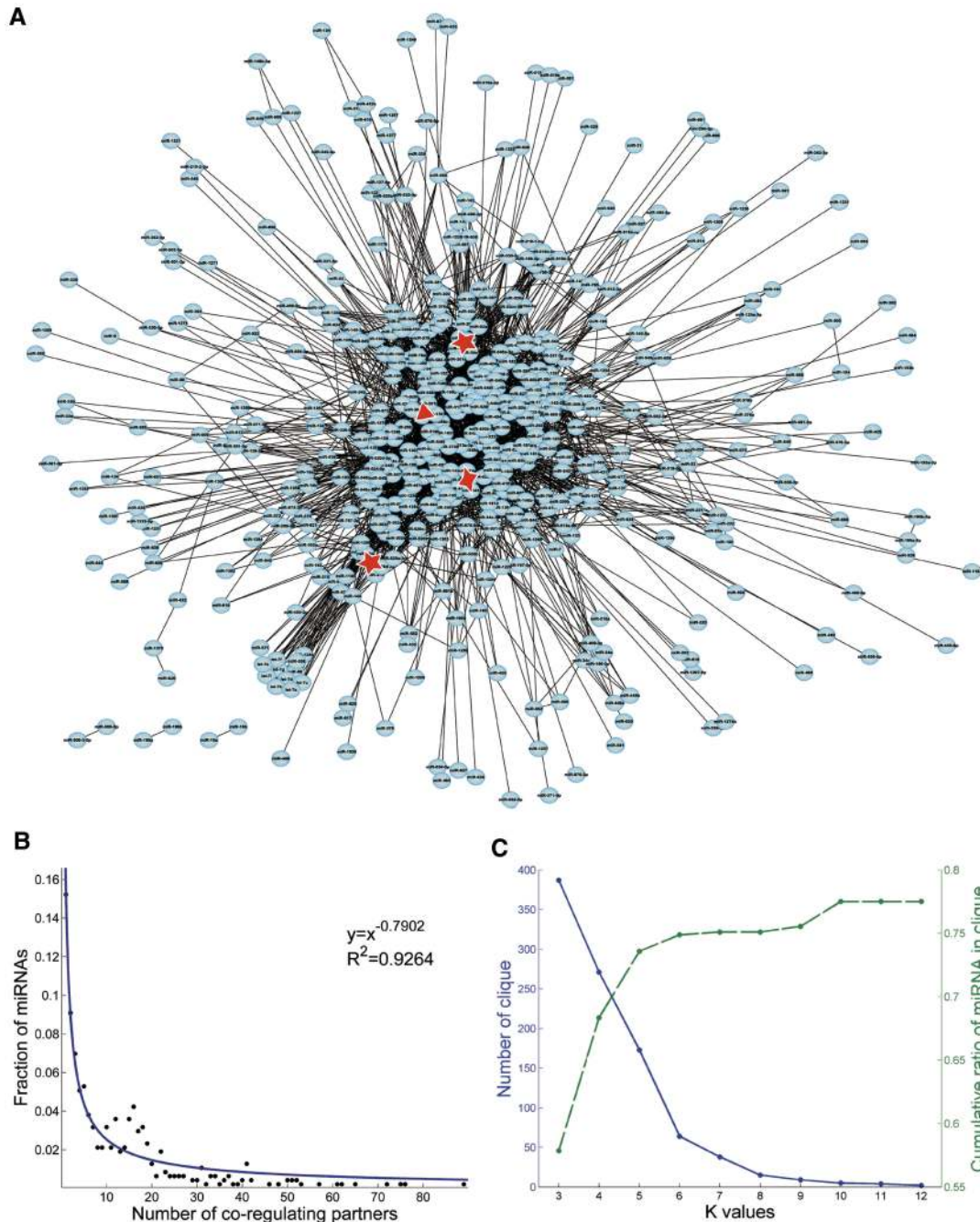
Figure 2. The layout of the MFSN and its structural features. (**A**) The MFSN generated by the procedure is described in the 'Materials and Methods' section. This network consists of 473 miRNAs and 2937 co-regulatory links. A node represents a miRNA, and an edge represents a synergistic action. An diamond marks the location of miRNAs associated with epithelial ovarian cancer, and a triangle marks the location of those associated with type 2 diabetes in Figure 3B. Pentacles mark the location of communities in Figure 4 with three *k*-values. (**B**) Degree distribution of the MFSN. (**C**) Number of cliques at different *k*-values and cumulative ratios of miRNAs in cliques with *k*-values are not bigger than *k*. The left *y*-axis represents number of cliques under different *k*-values, corresponding to the solid line. The right *y*-axis represents cumulative ratios of miRNAs in cliques, corresponding to the dashed line.

distribution of the MFSN reveals a power law with a slope of -0.7902 and $R^2 = \sim 0.9264$, showing that the MFSN is scale free and extending the result of Reut Shalgi *et al.* (14), who identified miRNA cooperation among 64 miRNAs. In addition, we found that nearly all miRNAs are connected together and the MFSN has a short diameter of 2.8691, which is similar to that of random

graphs generated by the duplication model (2.8722 ± 0.1332), as expected for a small-world network (43,44). The topology of the MFSN also exhibits dense local neighborhoods with an average clustering coefficient of 0.2747, which is much higher than for random networks (0.0684 ± 0.0151). This is because the immediate neighbors of a miRNA, its functional synergistic partners,

tend to be synergistic. The dense neighborhood feature of small-world networks is particularly interesting, because it can be exploited to predict synergism, as has been shown previously for protein–protein interactions (45).

To investigate the expression pattern of connecting miRNA pairs, we calculated their correlation coefficients and found that 76.80% of miRNA pairs have positive co-expression values (Supplementary Figure S1). Therefore, we proposed that most miRNA pairs with synergistic regulations tend to be co-expressed in all or most tissues studied, indicating that synergism is possible under most conditions. Meanwhile, only a small part of miRNA pairs co-expresses in tissue-restricted patterns, implying that their synergism might be under specific conditions. We concluded that the same expression tendency might ensure that miRNAs can promptly implement regulation under specific conditions to allow organisms to quickly adapt to a new environment.

Next, we analyzed the modular and community structure of the MFSN. Here, we defined a miRNA functional synergistic module as a clique that is a maximal complete subgraph. All modules and communities in the MFSN are identified using Cfinder (40). Each module (or community) has a unique composition of miRNAs, allowing the same miRNAs or the same pairs to occur in more than one module. Figure 2C shows the number of modules corresponding to every k-value, and the cumulative fraction of miRNAs found in modules. With an increase in the value of k, there is a sharp decrease in the number of modules. In total, 77.51% miRNAs are involved in at least one module. We interpreted this feature as a consequence that miRNAs implement specific regulation as small clusters rather than as individual or big modules. Because miRNAs from the same family tend to have similar functions or be involved in the same disease, we further investigated whether miRNAs of the same miRNA family occur in at least one module or community. Of the 70 miRNA families, 60% containing not fewer than two miRNAs are completely contained in at least one module, and the score for communities is larger (65.71%). Therefore, miRNAs from the same family do tend to be functional synergistic.

In all, the MFSN shows some generic properties: most miRNAs are connected and comprise a large sub network, and the network is scale free, modularity and small world.

MiRNAs for the same disease are close proximity in the MFSN

From the miR2Disease database, we obtained a total of 236 miRNAs involved in 108 diseases using the 'all disease data', and a total of 164 miRNAs correlated to 94 diseases are found using the 'high confidence disease data'. We further found that 75.93% diseases identified in the 'all disease data' involve at least two miRNAs; the value is 69.15% diseases in the 'high confidence disease data'. Because miRNAs associated with the same disease regulate similar or the same functional genes, they have synergistic actions. Next, we discussed whether they are close in the MFSN. Th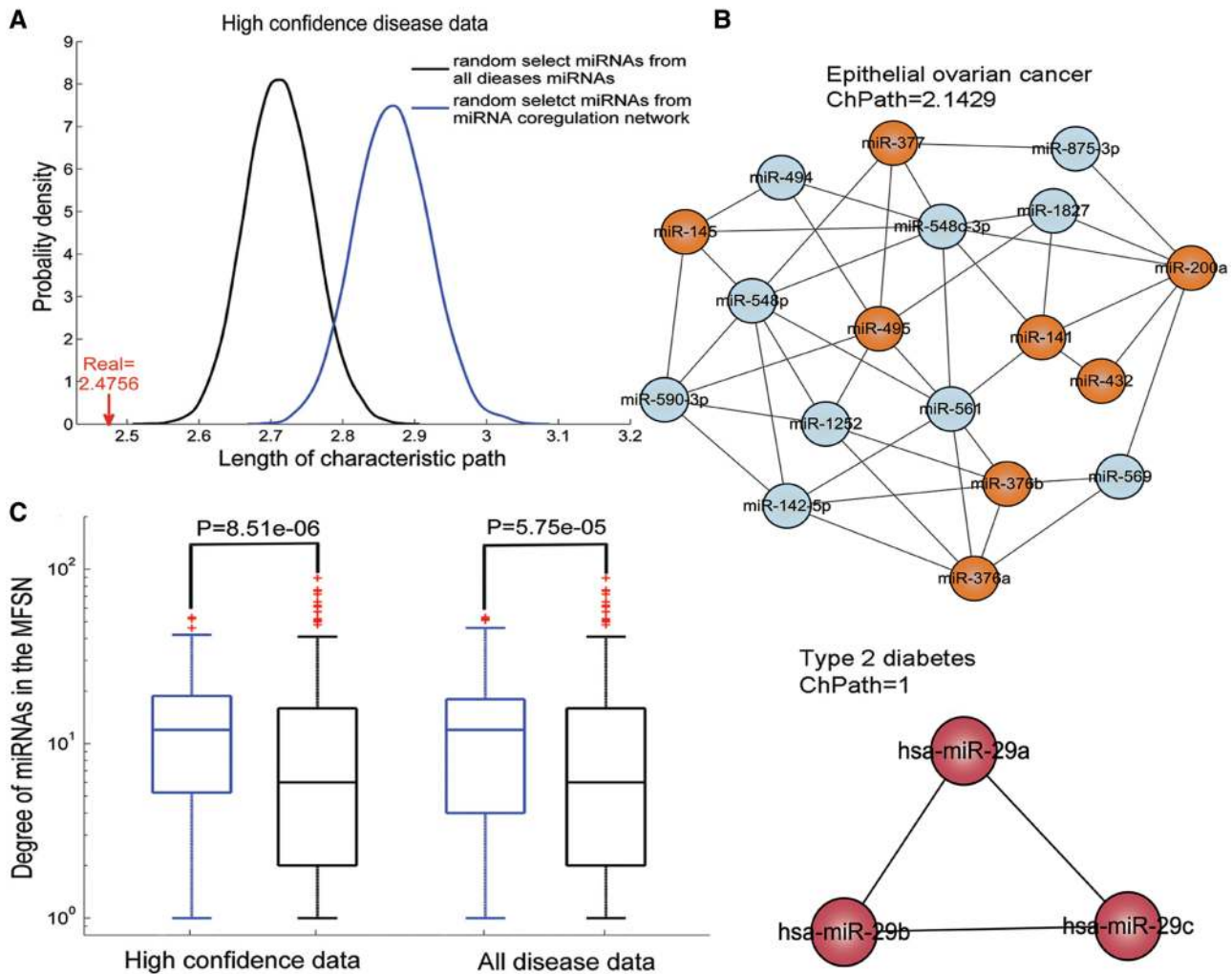e measure of characteristic path length among miRNAs for the same disease is used to assess the communication efficiency in the MFSN, which represents closeness and consequently how quickly information can spread in a network. The lower this measure is, the quicker the communication in the miRNA subsets is. Comparing the two classes of control-miRNA sets, we found that the characteristic path length of the same disease miRNAs is significantly lower (two P-values <0.00001, see 'Methods' section), indicating that miRNAs for the same disease are closer to each other in the MFSN. This tendency also exists in 'all disease data' (Supplementary Figure S2). So, we determined that miRNAs for the same disease tend to have direct or indirect, but not distant, functional synergy. For example, 8 of 14 miRNAs involved in epithelial ovarian cancer testified by low-throughput methods (46–49) have at least one synergistic partner and the characteristic path length among them is 2.3929, indicating that most miRNAs have direct functional synergism or share some partners (upper part of Figure 3B). Another example is the miR-29abc family, which contains hsa-miR-29a, hsa-miR-29b and hsa-miR-29c. A miRNA family incorporates similar mature miRNA sequences and complete identical seed regions, which are widely accepted as the 'key' regions for miRNA target identification (50). Therefore, their members are expected to have functional similarity and similar impacts. The three miRNAs are shown to have direct functional synergistic regulations (lower part of Figure 3B), and their expression levels have been found to be upregulated in type 2 diabetes, leading to insulin resistance in 3T3-L1 adipocytes (51).

As discussed above, miRNAs for the same disease are in close proximity in the MFSN to allow them to influence the same or similar functions by synergistic regulation of targets. It also indicates that miRNAs involved in the same disease have functional synergism, which is the basis of predicting new disease miRNAs.

Disease miRNA have more functional synergism

Degree is one of the most important topological measurements of a network and indicates local centrality of nodes in the network. Generally, the greater the degree, the more important is the node for the stabilization of the network. Therefore, we investigated whether disease miRNAs have a specific degree pattern in the MFSN. We divided miRNAs into two groups: disease miRNAs and non-disease miRNAs. Then we calculated the significance of the difference between the two groups and found that disease miRNAs have significantly higher degrees in the MFSN than non-disease miRNAs (Figure 3C); the median degree of disease miRNAs is 12 and that of non-disease miRNAs is 6. We obtained the same result using the 'all disease data'. Interestingly, we also found that the clustering coefficient of disease miRNAs is significantly larger than non-disease miRNAs (Supplementary Table S1). We proposed that it might be due to the features of miRNA family, the dysregulation of that would cause a similar phenotype (52).

The discovery of the difference in the number of synergistic partners between disease miRNAs and non-disease

Figure 3. Distinct topological features of disease miRNAs and two examples of diseases. (**A**) The mean characteristic path length among miRNAs for the same disease is shorter than both kinds of randomization tests. The arrow represents the mean characteristic path length in the real network, the line of light color is fitted using random selecting miRNAs from disease miRNAs and the line of dark color presents all miRNAs in the MFSN. (**B**) Two examples of characteristic path lengths (Chpath) of diseases. In the upper panel, nodes of dark color represent epithelial ovarian cancer-associated miRNAs; the inset panel shows location of the diamond in Figure 2A. The lower panel shows miRNAs associated with type 2 diabetes, and nodes of dark color represent associated miRNAs; the inset panel is located at the triangle in Figure 2A. (**C**) The difference in degrees between disease miRNAs and non-disease miRNAs with two types of data. Boxes of light color represent the distribution of disease miRNA degrees, and the black boxes correspond to non-disease miRNAs. *P*-values are calculated using the Wilcoxon rank-sum test.

miRNAs could suggest a difference in the functional complexity of these two groups. We measured the miRNA functional complexity by calculating the number of synergistically regulating function modules. As a result, the correlation between degree and functional complexity are found to be significantly strong positive ($R = 0.9055$, $P = 1.657e-177$), indicating that the more functional synergism miRNAs participate in, the more biological processes they regulate. Therefore, the dysregulation of miRNAs with more synergism would cause diseases. We also found that the trend is clearer in the 'high confidence disease data' than in the 'all disease data', suggesting that the false positive information incorporated in the 'all disease data' leads to a lower positive correlation (data not shown). We determined that the tendency would be stronger along with the increasing reliability and coverage of disease data.

Disease miRNAs are located at the interface of communities with high *k*-value

As discussed above, we knew that most miRNAs implement regulations as small modules and that miRNAs involved in the same disease are located in close proximity. Therefore, we investigated the modular features of disease miRNAs without distinguishing classes of diseases. As indicated in the Methods, a *k*-community comprises adjacent *k*-cliques, so a low *k*-value generates a large number of extensive communities of less tightly connected miRNA communities, showing a high degree of overlap, whereas increasing the *k*-value leads to fewer and more distinct miRNA communities that have a high degree of interconnection (Table 1). Interestingly, although cluster sizes decrease with increasing *k*-value, the proportion of disease miRNAs identified in the miRNA communities

increases, indicating the enrichment of disease miRNAs in the most tightly connected communities (Table 1). We could get the same tendency using the 'all disease data'. Here, we defined functions of a community using the functions that most miRNAs in the community synergistically regulate; for example, in the case of communities ($k = 11$), we found that the communities participate in different or similar biological processes (Supplementary Table S2 and Figure S3).

We further proposed that miRNAs as members of more than one community are of particular interest and more important, because miRNAs in multiple communities can be considered to be at the 'interface' of adjacent biological processes. 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 Table 2. When the k-value is 10, there are seven communities and 37.04% of disease miRNAs associated with these communities are located at the interface of communities (middle panel of Figure 4), which is 1.7284 times the ratio of non-disease miRNAs. In addition, five communities are identified when the k-value is 11, and 32% disease miRNAs are located at the interface, which is 2.1333 times more than with non-disease miRNAs (right panel of Figure 4). Therefore, disease miRNAs tend to be located at the interface, which can be considered as the interface of multiple functions. The locality can further classify the topological roles of miRNAs. Communities allow a degree of overlap, so we could distinguish the central importance of miRNAs, and a clear difference has been found between cancer protein and non-cancer protein (53). So, miRNAs in overlapping communities can be classified as global central cores and non-overlapping ones as local central cores. In all, most

disease miRNAs are located in the overlap region of most tightly connected communities. The above results thus highlight the key roles of disease miRNAs, which are reflected in their topological features in the MFSN.

DISCUSSION

In this study, we constructed the MFSN via co-regulating functional modules, which allows for an in-depth analysis of individual miRNAs in the context of their synergistic surroundings. As general biological networks, the MFSN is scale free, modular and has a small-world property. Watts and Strogatz (54) analyzed how fast disturbances spread through small-world networks and revealed that the time wasted for spreading of a disturbance in a small-world network is close to the theoretically possible minimum for any graph with the same number of nodes and edges. Therefore, small-worldness may allow the synergism of miRNAs to respond quickly to disturbances. Most synergistic miRNA pairs also have the same expression tendency, which allows for a rapid response to disturbances.

We not only identified miRNA synergism but also revealed their underlying functional patterns. Functional modules have three features at different levels, including common targets of miRNA pairs, significantly enriched GO categories and close proximity in the protein interaction network. In this study, a two-stage design is adopted to detect miRNA pairs with significant functional synergism, which are recently proposed to stepwise control the overall false discoveries (55,56). In the designs, a subset of miRNA pairs and their functional modules that pass functional enrichment significance threshold are chosen in the first stage, and in the second stage two topological restrictions in the protein interaction

Figure 4. Communities with k-values of 9, 10 and 11 show the tendency of location of disease miRNAs that are located at the pentacles in Figure 2A. Nodes of dark color represent disease miRNAs; miRNAs on the background occur in at least two communities.

Table 1. 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	9	231	124	34.93	203	152	42.82
4	17	170	112	39.72	150	132	46.81
5	21	127	92	42.01	112	107	48.86
6	22	91	75	45.18	78	88	53.01
7	19	67	63	48.46	57	73	56.15
8	9	50	43	46.24	42	51	54.84
9	7	34	35	50.72	27	42	60.87
10	7	28	27	49.09	21	34	61.82
11	5	20	25	55.56	16	29	64.44
12	2	5	19	79.17	4	20	83.33

The number of miRNA communities in the entire MFSN, identified by *k*-clique analysis at different *k*-values. Non_D_miRNA represents the number of non-disease miRNAs, D_miRNA is the number of disease miRNAs, D_miRNA_ratio is the fraction of disease miRNAs at different *k*-values.

Table 2. Multiple community membership distribution

k-value	High confidence disease data			All disease data		
	Observed (%)	Expected (%)	Fold differ	Observed (%)	Expected (%)	Fold differ
3	1.61	5.63	0.2866	2.63	5.42	0.4856
4	14.29	15.88	0.8995	15.91	14.67	1.0847
5	19.57	32.28	0.6060	20.56	33.04	0.6224
6	24.00	42.86	0.5600	25.00	44.87	0.5571
7	30.16	47.76	0.6314	34.25	45.61	0.7508
8	27.91	12.00	2.3256	27.45	9.52	2.8824
9	25.71	8.82	2.9143	23.81	7.41	3.2143
10	37.04	21.43	1.7284	41.18	9.52	4.3235
11	32.00	15.00	2.1333	37.93	0	Inf
12	0	0		0	0	

Percentage of disease miRNAs belonging to more than one community (based on miRNAs identified by clustering as belonging to a community). Expected value is based on non-disease miRNAs.

network are performed on this 'filtered' subset. In addition, when we controlled the false discovery rate (FDR < 0.12) (57,58) at the step of functional enrichment and kept other processes the same, we found that miRNA synergism identified by the two-stage design are included in the results using corrections of multiple testing. Therefore, these miRNA synergies are reliable.

The close proximity of miRNAs involved in the same disease is a further indication of their functional synergism and the complexity of disease. We have also shown that miRNAs involved in the disease exhibit a network topology different from that of miRNAs not yet identified as associated with a disease. The most striking property of disease miRNAs is their increased frequency of synergism. This observation indicates an underlying functional complexity to disease miRNAs. The *k*-clique clustering algorithm allows us to investigate miRNA synergism in a more informative way than just by looking at the interaction frequency of each miRNA. Its feature of overlapping communities allows us to distinguish between central and peripheral roles of miRNAs. The fact that we observed an enrichment of disease miRNAs at the

interface indicates their central roles. These results bridge the gap between the mechanism of disease miRNAs and the synergism among them.

In the text, we only analyzed the predicted highly efficient miRNA targets from TargetScan. To determine the effects of different target prediction algorithms, we further constructed another two MFSNs using miRNA targets from the highly efficient integrated data and miRBase database, respectively. These two MFSNs also provide similar power-law distributions, have small-world properties and most miRNAs also tend to work together as small clusters. More importantly, topological features of disease miRNAs are distinct from those of non-disease miRNAs; these results are accordant with those obtained using TargetScan. MiRNAs associated with the same disease are closely located in the MFSNs. Meanwhile, disease miRNAs have significantly more synergistic partners than non-disease miRNAs, reside in communities with high *k* values and tend to be located at the interface of communities. The detailed results obtained with the integrated data and miRBase are shown in the Supplementary Data and further evidence that the structures of the MFSNs and the topological features of disease miRNAs are robust.

The results present here provide a new insight into the global topological properties of disease miRNAs in the comprehensive microRNA synergistic network. Although limitations exist in the current data, the results uncovered here are important for understanding the key roles of miRNAs in diseases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

REFERENCES

1. Farh,K.K., Grimson,A., Jan,C., Lewis,B.P., Johnston,W.K., Lim,L.P., Burge,C.B. and Bartel,D.P. (2005) The widespread impact of mammalian MicroRNAs on mRNA repression and evolution. *Science*, **310**, 1817–1821.
2. Friedman,J.M. and Jones,P.A. (2009) MicroRNAs: critical mediators of differentiation, development and disease. *Swiss Med. Wkly*, **139**, 466–472.
3. Migliore,C. and Giordano,S. (2009) MiRNAs as new master players. *Cell Cycle*, **8**, 2185–2186.
4. Bentwich,I., Avniel,A., Karov,Y., Aharonov,R., Gilad,S., Barad,O., Barzilai,A., Einat,P., Einav,U., Meiri,E. *et al.* (2005) Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.*, **37**, 766–770.
5. 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.
6. Friedman,R.C., Farh,K.K., Burge,C.B. and Bartel,D.P. (2009) Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.*, **19**, 92–105.
7. Backes,C., Meese,E., Lenhof,H.P. and Keller,A. (2010) A dictionary on microRNAs and their putative target pathways. *Nucleic Acids Res.*, **38**, 4476–4486.
8. Enright,A.J., John,B., Gaul,U., Tuschl,T., Sander,C. and Marks,D.S. (2003) MicroRNA targets in Drosophila. *Genome Biol.*, **5**, R1.
9. 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.
10. Wu,S., Huang,S., Ding,J., Zhao,Y., Liang,L., Liu,T., Zhan,R. and He,X. Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3′ untranslated region. *Oncogene*, **29**, 2302–2308.
11. Shalgi,R., Lieber,D., Oren,M. and Pilpel,Y. (2007) Global and local architecture of the mammalian microRNA-transcription factor regulatory network. *PLoS Comput. Biol.*, **3**, e131.
12. Boross,G., Orosz,K. and Farkas,I.J. (2009) Human microRNAs co-silence in well-separated groups and have different predicted essentialities. *Bioinformatics*, **25**, 1063–1069.
13. Zhou,Y., Ferguson,J., Chang,J.T. and Kluger,Y. (2007) Inter- and intra-combinatorial regulation by transcription factors and microRNAs. *BMC Genomics*, **8**, 396.
14. An,J., Choi,K.P., Wells,C.A. and Chen,Y.P. Identifying co-regulating microrna groups. *J. Bioinform. Comput. Biol.*, **8**, 99–115.
15. Papadopoulos,G.L., Alexiou,P., Maragkakis,M., Reczko,M. and Hatzigeorgiou,A.G. (2009) DIANA-mirPath: integrating human and mouse microRNAs in pathways. *Bioinformatics*, **25**, 1991–1993.
16. Liang,H. and Li,W.H. (2007) MicroRNA regulation of human protein protein interaction network. *RNA*, **13**, 1402–1408.
17. Yuan,X., Liu,C., Yang,P., He,S., Liao,Q., Kang,S. and Zhao,Y. (2009) Clustered microRNAs' coordination in regulating protein-protein interaction network. *BMC Syst. Biol.*, **3**, 65.
18. Makeyev,E.V. and Maniatis,T. (2008) Multilevel regulation of gene expression by microRNAs. *Science*, **319**, 1789–1790.
19. Luo,X., Lin,H., Pan,Z., Xiao,J., Zhang,Y., Lu,Y., Yang,B. and Wang,Z. (2008) Down-regulation of miR-1/miR-133 contributes to re-expression of pacemaker channel genes HCN2 and HCN4 in hypertrophic heart. *J. Biol. Chem.*, **283**, 20045–20052.
20. Baek,D., Villen,J., Shin,C., Camargo,F.D., Gygi,S.P. and Bartel,D.P. (2008) The impact of microRNAs on protein output. *Nature*, **455**, 64–71.
21. Selbach,M., Schwanhausser,B., Thierfelder,N., Fang,Z., Khanin,R. and Rajewsky,N. (2008) Widespread changes in protein synthesis induced by microRNAs. *Nature*, **455**, 58–63.
22. Griffiths-Jones,S., Saini,H.K., van Dongen,S. and Enright,A.J. (2008) miRBase: tools for microRNA genomics. *Nucleic Acids Res.*, **36**, D154–D158.
23. 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–W276.
24. John,B., Enright,A.J., Aravin,A., Tuschl,T., Sander,C. and Marks,D.S. (2004) Human MicroRNA targets. *PLoS Biol.*, **2**, e363.
25. 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.
26. Rehmsmeier,M., Steffen,P., Hochsmann,M. and Giegerich,R. (2004) Fast and effective prediction of microRNA/target duplexes. *RNA*, **10**, 1507–1517.
27. Keshava Prasad,T.S., Goel,R., Kandasamy,K., Keerthikumar,S., Kumar,S., Mathivanan,S., Telikicherla,D., Raju,R., Shafreen,B., Venugopal,A. *et al.* (2009) Human Protein Reference Database— 2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
28. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
29. Martin,D., Brun,C., Remy,E., Mouren,P., Thieffry,D. and Jacq,B. (2004) GOToolBox: functional analysis of gene datasets based on Gene Ontology. *Genome Biol.*, **5**, R101.
30. Mintseris,J. and Weng,Z. (2005) Structure, function, and evolution of transient and obligate protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **102**, 10930–10935.
31. Shankavaram,U.T., Varma,S., Kane,D., Sunshine,M., Chary,K.K., Reinhold,W.C., Pommier,Y. and Weinstein,J.N. (2009) CellMiner: a relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics*, **10**, 277.
32. Shankavaram,U.T., Reinhold,W.C., Nishizuka,S., Major,S., Morita,D., Chary,K.K., Reimers,M.A., Scherf,U., Kahn,A., Dolginow,D. *et al.* (2007) Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol. Cancer Ther.*, **6**, 820–832.
33. Wu,Z., Irizarry,R.A., Gentleman,R., Martinez-Murillo,F. and Spencer,F. (2004) A model-based background adjustment for oligonucleotide expression arrays. *J. Am. Stat. Assoc.*, **99**, 909–917.
34. Blower,P.E., Verducci,J.S., Lin,S., Zhou,J., Chung,J.H., Dai,Z., Liu,C.G., Reinhold,W., Lorenzi,P.L., Kaldjian,E.P. *et al.* (2007) MicroRNA expression profiles for the NCI-60 cancer cell panel. *Mol. Cancer Ther.*, **6**, 1483–1491.
35. Jiang,Q., Wang,Y., Hao,Y., Juan,L., Teng,M., Zhang,X., Li,M., Wang,G. and Liu,Y. (2009) miR2Disease: a manually curated database for microRNA deregulation in human disease. *Nucleic Acids Res.*, **37**, D98–D104.
36. Milo,R., Kashtan,N., Itzkovitz,S., Newman,M.E.J. and Alon,U. (2003) Uniform generation of random networks with arbitrary degree sequence. *condmat/0312028*, Retrieved from http://arxiv.org/abs/cond-mat/0312028v1.
37. Barabasi,A.L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.*, **5**, 101–113.
38. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
39. Palla,G., Derenyi,I., Farkas,I. and Vicsek,T. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, **435**, 814–818.
40. Adamcsek,B., Palla,G., Farkas,I.J., Derenyi,I. and Vicsek,T. (2006) CFinder: locating cliques and overlapping modules in biological networks. *Bioinformatics*, **22**, 1021–1023.

41. Hebert,S.S., Horre,K., Nicolai,L., Papadopoulou,A.S., Mandemakers,W., Silahtaroglu,A.N., Kauppinen,S., Delacourte,A. and De Strooper,B. (2008) Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. *Proc. Natl Acad. Sci. USA*, **105**, 6415–6420.

42. Sayed,D., Hong,C., Chen,I.Y., Lypowy,J. and Abdellatif,M. (2007) MicroRNAs play an essential role in the development of cardiac hypertrophy. *Circ. Res.*, **100**, 416–424.

43. Barabasi,A.L. and Albert,R. (1999) Emergence of scaling in random networks. *Science*, **286**, 509–512.

44. Barabasi,A.L. and Bonabeau,E. (2003) Scale-free networks. *Sci. Am.*, **288**, 60–69.

45. Goldberg,D.S. and Roth,F.P. (2003) Assessing experimentally derived interactions in a small world. *Proc. Natl Acad. Sci. USA*, **100**, 4372–4376.

46. Zhang,L., Volinia,S., Bonome,T., Calin,G.A., Greshock,J., Yang,N., Liu,C.G., Giannakakis,A., Alexiou,P., Hasegawa,K. *et al.* (2008) Genomic and epigenetic alterations deregulate microRNA expression in human epithelial ovarian cancer. *Proc. Natl Acad. Sci. USA*, **105**, 7004–7009.

47. Iorio,M.V., Visone,R., Di Leva,G., Donati,V., Petrocca,F., Casalini,P., Taccioli,C., Volinia,S., Liu,C.G., Alder,H. *et al.* (2007) MicroRNA signatures in human ovarian cancer. *Cancer Res.*, **67**, 8699–8707.

48. Chen,R., Alvero,A.B., Silasi,D.A., Kelly,M.G., Fest,S., Visintin,I., Leiser,A., Schwartz,P.E., Rutherford,T. and Mor,G. (2008) Regulation of IKKbeta by miR-199a affects NF-kappaB activity in ovarian cancer cells. *Oncogene*, **27**, 4712–4723.

49. Lu,L., Katsaros,D., de la Longrais,I.A., Sochirca,O. and Yu,H. (2007) Hypermethylation of let-7a-3 in epithelial ovarian cancer is associated with low insulin-like growth factor-II expression and favorable prognosis. *Cancer Res.*, **67**, 10117–10122.

50. Bartel,D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215–233.

51. He,A., Zhu,L., Gupta,N., Chang,Y. and Fang,F. (2007) Overexpression of microribonucleic acid 29, highly up-regulated in diabetic rats, leads to insulin resistance in 3T3-L1 adipocytes. *Mol. Endocrinol.*, **21**, 2785–2794.

52. Lu,M., Zhang,Q., Deng,M., Miao,J., Guo,Y., Gao,W. and Cui,Q. (2008) An analysis of human microRNA and disease associations. *PLoS ONE*, **3**, e3420.

53. Jonsson,P.F. and Bates,P.A. (2006) Global topological features of cancer proteins in the human interactome. *Bioinformatics*, **22**, 2291–2297.

54. Watts,D.J. and Strogatz,S.H. (1998) Collective dynamics of 'small-world' networks. *Nature*, **393**, 440–442.

55. Lin,D.Y. (2006) Evaluating statistical significance in two-stage genomewide association studies. *Am. J. Hum. Genet.*, **78**, 505–509.

56. McCarthy,M.I., Abecasis,G.R., Cardon,L.R., Goldstein,D.B., Little,J., Ioannidis,J.P. and Hirschhorn,J.N. (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.*, **9**, 356–369.

57. Hochberg,Y.B.a.Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Stat. Soc.*, **57**, 289–300.

58. Ule,J., Ule,A., Spencer,J., Williams,A., Hu,J.S., Cline,M., Wang,H., Clark,T., Fraser,C., Ruggiu,M. *et al.* (2005) Nova regulates brain-specific splicing to shape the synapse. *Nat. Genet.*, **37**, 844–852.