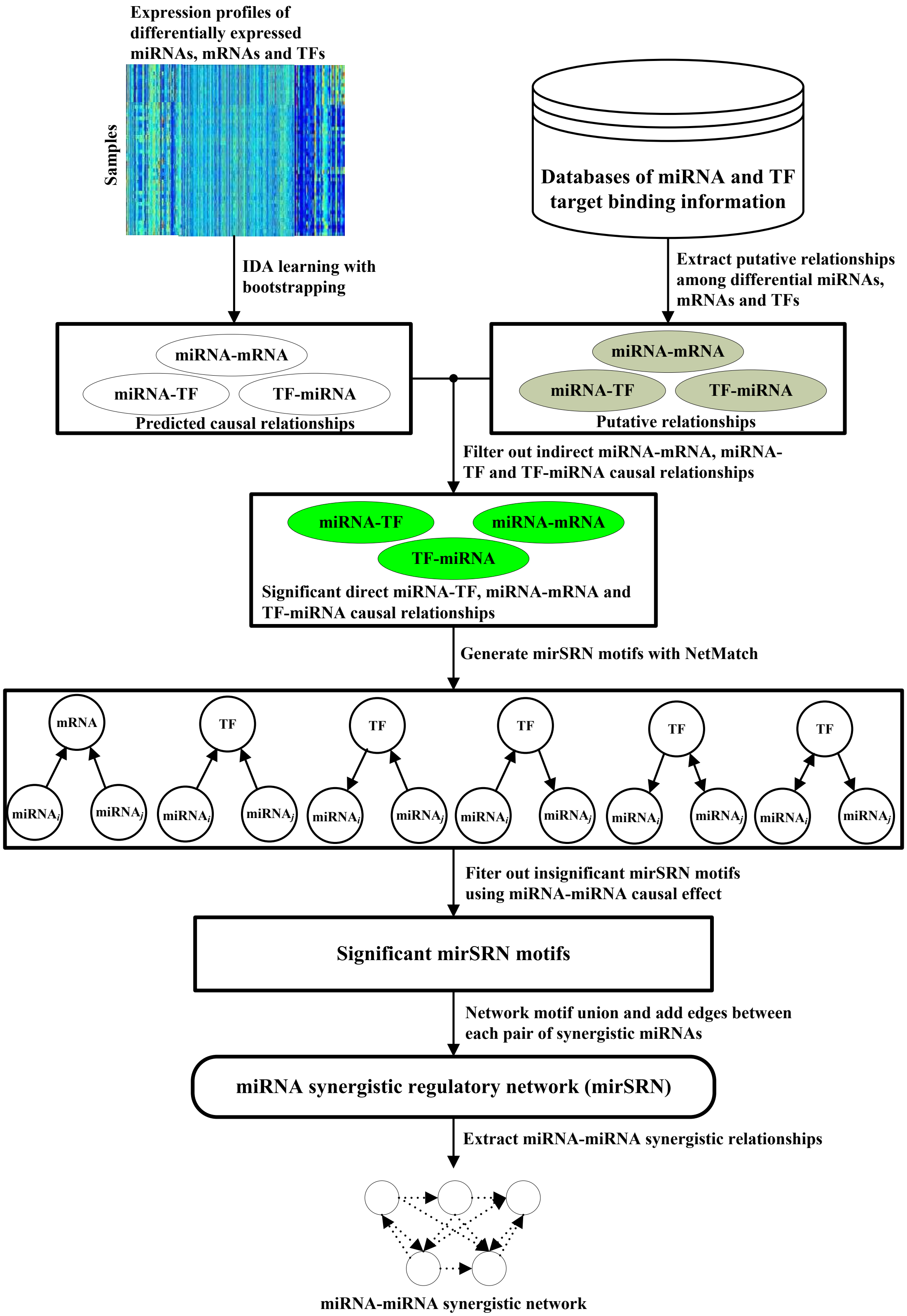

Expression profiles of differentially expressed miRNAs, mRNAs and TFs
Samples
Databases of miRNA and TF target binding information
IDA learning with bootstrapping
Extract putative relationships among differential miRNAs, mRNAs and TFs
miRNA-mRNA
miRNA-TF
TF-miRNA
Predicted causal relationships
miRNA-mRNA
miRNA-TF
TF-miRNA
Putative relationships
Filter out indirect miRNA-mRNA, miRNA-TF and TF-miRNA causal relationships
miRNA-TF
miRNA-mRNA
TF-miRNA
Significant direct miRNA-TF, miRNA-mRNA and TF-miRNA causal relationships
Generate mirSRN motifs with NetMatch
mRNA
TF
TF
TF
TF
TF
miRNA$_i$
miRNA$_j$
miRNA$_i$
miRNA$_j$
miRNA$_i$
miRNA$_j$
miRNA$_i$
miRNA$_j$
miRNA$_i$
miRNA$_j$
miRNA$_i$
miRNA$_j$
Fiter out insignificant mirSRN motifs using miRNA-miRNA causal effect
Significant mirSRN motifs
Network motif union and add edges between each pair of synergistic miRNAs
miRNA synergistic regulatory network (mirSRN)
Extract miRNA-miRNA synergistic relationships
miRNA-miRNA synergistic network

Journal Name

ARTICLE

Identifying miRNA synergistic regulatory network in heterogeneous human data via network motifs†

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/molecularbiosystems

Junpeng Zhang,[a,*] Thuc D. Le,[b] Lin Liu,[b] Jianfeng He[c] and Jiuyong Li[b,*]

Understanding the synergism of multiple microRNAs (miRNAs) in gene regulation can provide important insight into the mechanisms of complex human diseases caused by miRNA regulation. Therefore, it is important to identify miRNA synergism and study miRNA characteristics in miRNA synergistic regulatory networks. A number of methods have been proposed to identify miRNA synergism. However, most of the methods only use downstream target genes of miRNAs to infer miRNA synergism when miRNAs can also be regulated by upstream transcription factors (TFs) at the transcriptional level. Additionally, most methods are based on statistical associations identified from data without considering the causal nature of gene regulation. In this paper, we present a causality based framework, called mirSRN (miRNA Synergistic Regulatory Network) to infer miRNA synergism in human molecular systems by considering both downstream miRNA targets and upstream TF regulation. We apply the proposed framework to two real world datasets and discover that almost all the top 10 miRNAs with the largest node degree in the mirSRNs are associated with different human diseases, including cancers, and that the mirSRNs are approximately scale-free and small-world networks. We also find that most miRNAs in the networks are frequently synergistic with other miRNAs, and miRNAs related to the same disease are likely to be synergistic and in a cluster linking to a biological function. Synergistic miRNA pairs show higher co-expression level, and may have potential functional relationships indicating collaboration between the miRNAs. Functional validation of the identified synergistic miRNAs demonstrates that these miRNAs cause different kinds of diseases. These results deepen our understanding of the biological meaning of miRNA synergism.

1 Introduction

MicroRNAs (miRNAs) are small (~23nt), non-coding RNAs found in plants and animals, and they mainly regulate gene expression at the post-transcriptional level.[1] By binding with complementary sequences within mRNA (messenger RNA) molecules, miRNAs may result in gene silencing via the full degradation of the target mRNA transcript or translational repression of it.[2] About 60% of genes are predicted to be regulated by miRNAs in human genome.[3, 4] Given their far-reaching role, it is not surprising that miRNAs are likely to be involved in most biological processes, including developmental timing, cell proliferation, metabolism, differentiation, apoptosis, cellular signalling, stress responses and even cancer.[5-11]

Additionally, combinatorial regulation is a feature of miRNA regulation. A miRNA can target multiple mRNAs, and one mRNA can be targeted by multiple miRNAs.[12-14] Recent studies show that complex diseases are affected by several miRNAs rather than a single miRNA. Therefore, it is important to identify miRNA synergism to determine miRNA functions at a system-wide level and investigate disease miRNA features in miRNA synergistic regulatory networks.

With the availability of large volume of data obtained from high-throughput experiments and the advance of computer algorithms and tools,[15-19] it becomes possible to investigate the complex synergistic relationships between miRNAs using computational approaches. Our understanding of miRNA synergistic regulation is increasing with help of the computational methods developed for uncovering miRNA synergism. The computational methods can be classified into two main streams: (1) those using sequence-based target binding information alone[20-26] and (2) methods combining expression data with putative sequence-based target binding information.[27-33] The sequence-based target binding information mainly includes predicted miRNA targets, protein-protein interaction (PPI) and miRNA-SNP interactions, and the expression data contains expression profiles of miRNAs and mRNAs. They construct miRNA-miRNA synergistic network based on two main constraints: (1) Significant overlap of target genes between each miRNA-miRNA pair using hypergeometric distribution test, and (2) Overlap of target genes between each miRNA-miRNA pair significantly enriched Gene Ontology (http://geneontology.org/, GO) terms or Kyoto Encyclopedia of Genes and Genomes (http://www.genome.jp/kegg/, KEGG) pathways.

The above approaches[20-33] discover miRNA clusters or regulatory modules using statistical tests, thus the identified

[a] School of Engineering, Dali University, Dali, Yunnan, 671003, P. R. China. E-mail: zhangjunpeng_411@yahoo.com
[b] School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, SA 5095, Australia. E-mail: jiuyong.li@unisa.edu.au
[c] Institute of Biomedical Engineering, Kunming University of Science and Technology, Kunming, Yunnan, 650500, P. R. China.

† Electronic Supplementary Information (ESI) available: See DOI: 10.1039/x0xx00000x

relationships may not reveal gene regulatory relationships which are indeed causal relationships. Furthermore, most existing methods only use downstream target genes of miRNAs to study miRNA synergism. However, miRNAs can also be regulated by upstream TFs at the transcriptional level. Therefore, it is necessary to use both common target genes and regulators of miRNAs to infer miRNA synergism.

Based on the above observations, we propose a causal discovery based framework, mirSRN (miRNA synergistic regulatory network) to discover miRNA synergism by using heterogeneous human data, including matched miRNA and mRNA expression data, and putative target binding information of miRNA-mRNA, miRNA-TF and TF-miRNA. Our method different from the previous methods[20-33] identifies miRNA-miRNA synergistic network based on two different constraints: (1) Significant causal relationships of miRNA-target, TF-miRNA using causality based method in human heterogeneous data, and (2) Strength of each miRNA-miRNA pair measured by the miRNA-miRNA causal effects.

As illustrated in Fig. 1, TFs, miRNAs and mRNAs are regarded as upstream, midstream and downstream molecules, respectively. We suppose that multiple miRNAs can directly regulate the same target by cooperating with each other. We call such a regulation pattern a direct miRNA synergistic regulatory network (DmirSRN) motif. Meanwhile, miRNA synergism can be indirect, i.e. as shown in Fig. 1, the regulation by a miRNA (on its target) is influenced or regulated by a TF and the TF in turn is regulated by another miRNA. We call this pattern an indirect miRNA synergistic regulatory network (ImirSRN) motif. We have identified 2 DmirSRN motifs and 4 ImirSRN motifs as shown in Fig. 1. With our proposed framework we identify both types of motifs and combine them to build a miRNA synergistic network (we also call the network mirSRN in this paper).

We apply the proposed framework to two human datasets: the epithelial-mesenchymal transition (EMT) and multi class cancer (MCC) datasets respectively. We discover that the top 10 miRNAs with largest node degrees of the built mirSRNs are almost all associated with different human diseases of the EMT and MCC datasets, including cancers, and the mirSRNs are approximately scale-free and small-world networks. In addition, we extract from a mirSRN the sub-network that only contains the miRNA-miRNA synergistic relationships (called *miRNA-miRNA synergistic network* in this paper), and we have found that all miRNAs in each of the miRNA-miRNA synergistic networks are closely connected, resulting in a high percentage of *hub* miRNAs, which indicates that most miRNAs are frequently synergistic with other miRNA partners. We also find that the miRNAs related to the same disease are likely to be synergistic and form clusters (highly interconnected regions) in the miRNA-miRNA synergistic networks. Moreover, synergistic miRNA pairs show higher co-expression level than that of non-synergistic miRNA pairs. Finally, we demonstrate that the identified synergistic miRNAs can cause different kinds of diseases.

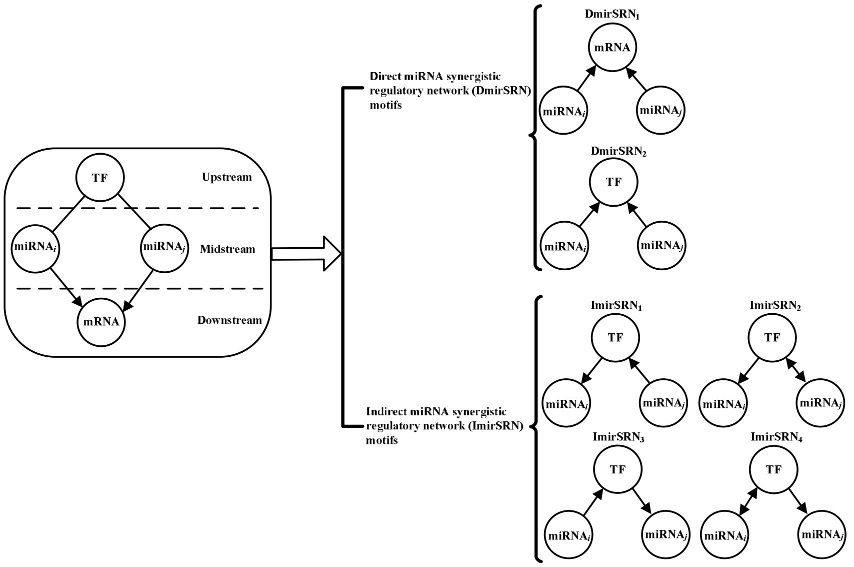


Fig. 1 TF, miRNA and mRNA are regarded as upstream, midstream and downstream molecules, respectively. There are two types of miRNA synergistic regulatory (mirSRN) motifs: direct (DmirSRN) and indirect (ImirSRN) motifs. The numbers of DmirSRN and ImirSRN motifs are 2 and 4, respectively.

2 Materials and methods

2.1 Data selection and processing

We use the matched miRNA and mRNA expression profiles from the epithelial-mesenchymal transition (EMT) and multi class cancer (MCC) datasets in this work. The miRNA expression profiles of EMT are from Søkilde *et al.*[34] They are profiled from the 60 cancer cell lines of the drug screening panel of human cancer cell lines at the National Cancer Institute (NCI-60). They are available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE26375. The mRNA expression profiles of EMT for NCI-60 are obtained from ArrayExpress (http://www.ebi.ac.uk/arrayexpress, accession number E-GEOD-5720). Samples of EMT categorized as epithelial (11 samples) and mesenchymal (36 samples) are used for this work. The MCC dataset is also composed of matched miRNA and mRNA expression data. The miRNA expression profiles of MCC are obtained from Lu *et al.*[35] They are available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE2564. The mRNA expression profiles of MCC are from Ramaswamy *et al*.[36] They can be downloaded at http://www.broad.mit.edu/cancer/pub/migcm. Samples of MCC classified as normal (21 samples) and tumor (67 samples) are used in this work. To extract all the TF genes in both the EMT and MCC datasets, we use the list of TF repertoire.[37] This list will then be used to query against the mRNA expression profiles to obtain TF expression profiles for both the EMT and MCC datasets.

Putative miRNA-target (miRNA-mRNA and miRNA-TF) binding information is obtained by using the union of three databases: MicroCosm (Version v5),[38] TargetScan (Version v7.0),[39] and starBase (Version v2.0).[40] The TF-miRNA target information is downloaded from MIR@NT@N (Version v1.2.1).[41] In this study, we are only interested in the putative target binding information of miRNA-mRNA, miRNA-TF and TF-miRNA for discovering miRNA synergistic regulatory networks.

2.2 Overview of the proposed framework

As shown in Fig. 2, the overall process of our framework for discovering a mirSRN comprises the steps as described in the following.

(1) Differentially expressed gene analysis. Given the matched miRNA and mRNA expression profiles, a list of differentially expressed miRNAs, mRNAs and TFs are identified. The expression profiles of the differentially expressed miRNAs, mRNAs and TFs are integrated into an input dataset or matrix of the next step.

(2) Inferring significant causal regulatory relationships. We firstly use the causal inference method, IDA (Intervention calculus when the DAG (Directed Acyclic Graph) is Absent),[42, 43] to calculate the causal (regulatory) effects of a miRNA or TF using the expression data. Considering the expression level of each miRNA, mRNA and TF as a random variable, the IDA learning contains two main steps: (a) learning the causal structure of the variables from expression data using the PC algorithm;[44] (b) based on the causal structure, infer the causal effect of one variable on the other using the do-calculus[45] and the expression data. Given the focus of this paper, we refer readers to the reference[46] for the details of the two steps when used in finding miRNA regulatory relationships.

For each miRNA, its causal effect on every mRNA and TF is calculated; and for each TF, its causal effect on every miRNA is calculated too. If the causal effect (absolute value) is larger than the median causal effect, we consider that there is a significant causal relationship between the pair of miRNA-mRNA, miRNA-TF, or TF-miRNA, and non-significant relationships are discarded. A significant causal relationship may not be a result of the direct interaction between the pair, e.g. the causal relationship between a miRNA and mRNA may be mediated by a TF. To obtain significant direct causal interactions, we filter out indirect causal relationships of miRNA-mRNA, miRNA-TF and TF-miRNA using the miRNA and TF putative target information retrieved from the databases described in the previous section. All the direct significant causal relationships are to be used to generate the network motifs in the next step.

(3) Identifying DmirSRN and ImirSRN motifs. The network motifs can be considered as simple building blocks from which the network is composed,[47] and are believed to have specific functions which play critical roles in biological network inference.[48] Furthermore, the identified network motifs could provide insights into the synergistic relationships between miRNAs. The significant direct causal interactions of miRNA-mRNA and miRNA-TF are used to identify the DmirSRN motifs with NetMatch.[49] Similarly, the direct miRNA-TF and TF-miRNA interactions obtained in the previous step are used to identify the ImirSRN motifs.

(4) Filtering out insignificant DmirSRN and ImirSRN motifs. As shown in Fig. 1, each pair of miRNAs contained in a DmirSRN or ImirSRN motif has a synergistic relationship as they either co-regulate the same target or one miRNA's regulation on its target is indirectly influenced by the other miRNA. However, the strength of the miRNA synergism varies, and we are only interested in strong miRNA synergism. Furthermore, as there may exist false discoveries in the miRNA-mRNA, miRNA-TF and TF-miRNA causal relationships identified in the previous step, the miRNA synergism deduced only from the motifs may have false discoveries too. Therefore we want to use the strength of the miRNA synergism to quantify and thus to filter out insignificant miRNA synergism to reduce the false discoveries. In our work, the strength of miRNA synergism is measured by the miRNA-miRNA causal effect calculated using IDA. If the causal effect (absolute value) of a miRNA-miRNA synergistic pair is larger than the median causal effect (of all the miRNA-miRNA causal effects), we consider that there is a strong or significant synergism between the pair of miRNAs. The DmirSRN and ImirSRN motifs with insignificant miRNA-miRNA synergistic pairs are discarded.

(5) Building mirSRN. After filtering out insignificant DmirSRN and ImirSRN motifs, the union of all the significant DmirSRN and ImirSRN motifs are taken to build mirSRN. Then we add edges between each pair of synergistic miRNAs. Note that the added edges between pairs of miRNAs indicate synergistic relationships rather than direct biological interactions, which is unlike the miRNA-TF, miRNA-mRNA, or TF-miRNA edges in the mirSRN.

(6) Creating miRNA-miRNA synergistic network. We extract from a mirSRN only the miRNA-miRNA relationships, which are the miRNA-miRNA synergistic relationships as indicated by the significant causal effects between the pairs of the miRNAs and more importantly by the different patterns of synergism of miRNAs in gene regulation as shown in Fig. 1. As noted earlier, in this paper, we call the network formed by the extracted miRNA-miRNA synergistic relationships the miRNA-miRNA synergistic network.

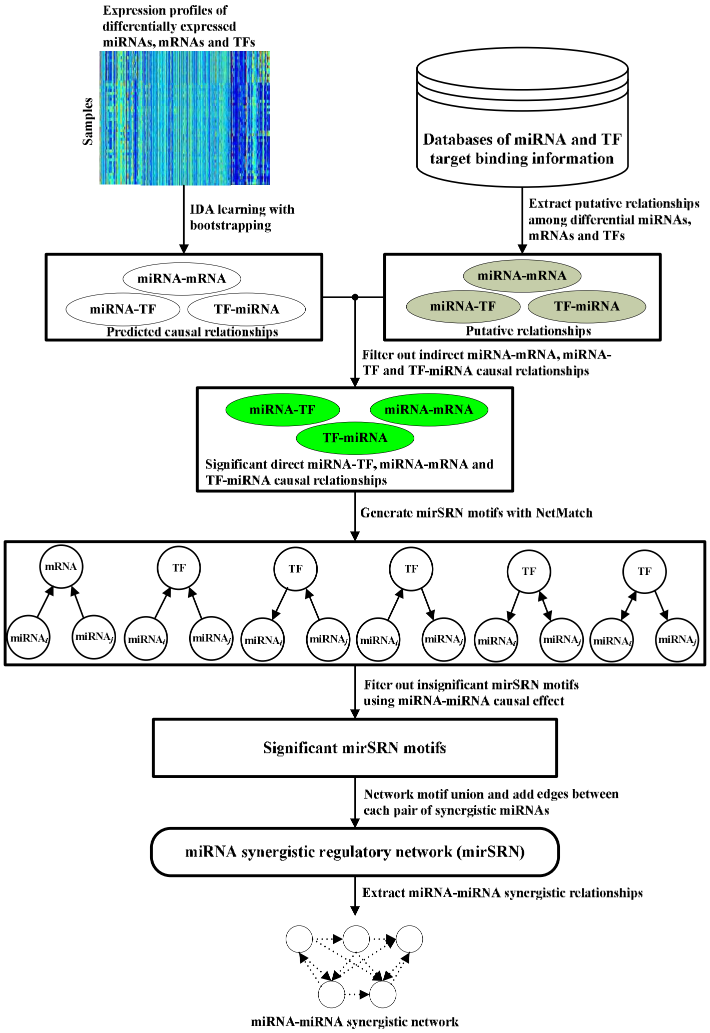

Expression profiles of differentially expressed miRNAs, mRNAs and TFs
Samples
Databases of miRNA and TF target binding information
IDA learning with bootstrapping
Extract putative relationships among differential miRNAs, mRNAs and TFs
miRNA-mRNA
miRNA-TF
TF-miRNA
Predicted causal relationships
miRNA-mRNA
miRNA-TF
TF-miRNA
Putative relationships
Filter out indirect miRNA-mRNA, miRNA-TF and TF-miRNA causal relationships
miRNA-TF
miRNA-mRNA
TF-miRNA
Significant direct miRNA-TF, miRNA-mRNA and TF-miRNA causal relationships
Generate mirSRN motifs with NetMatch
mRNA
TF
TF
TF
TF
TF
miRNAi
miRNAj
miRNAi
miRNAj
miRNAi
miRNAj
miRNAi
miRNAj
miRNAi
miRNAj
miRNAi
miRNAj
Fiter out insignificant mirSRN motifs using miRNA-miRNA causal effect
Significant mirSRN motifs
Network motif union and add edges between each pair of synergistic miRNAs
miRNA synergistic regulatory network (mirSRN)
Extract miRNA-miRNA synergistic relationships
miRNA-miRNA synergistic network

Fig. 2 A workflow of the proposed framework to construct mirSRN. The process involves three main steps. We firstly identify causal relationships with IDA learning, and generate significant causal relationships by combining putative target binding information. Then, we assemble significant causal relationships of miRNA-mRNA, miRNA-TF and TF-miRNA, and filter out insignificant mirSRN motifs using miRNA-miRNA causal effects to infer mirSRN. Finally, we extract miRNA-miRNA synergistic relationships to create miRNA-miRNA synergistic network. Since the causal effect from a miRNA *i* (*j*) to another miRNA *j* (*i*) is different, we use directed arrows with dotted lines to denote synergistic relationships between miRNAs.

2.3 Analysis of network topological properties

We analyze miRNA synergism by examining both mirSRNs and the miRNA-miRNA synergistic networks extracted from the mirSRNs. mirSRNs are used to understand the relationships between miRNAs, TFs and their targets, and miRNA-miRNA synergistic networks are used to understand functional synergistic relationships between miRNAs.

Network topological properties provide valuable insight into the internal organization of a network, e.g. power law degree distribution, clustering coefficient, the average clustering coefficient, characteristic path length and the average characteristic path length. If a network whose degree distribution follows a power law, the network is scale-free, and this is an important characteristic of real-world biological networks.[50] Clustering coefficient is a measure of the degree to which nodes in a network tend to cluster together, and the average clustering coefficient is used to evaluate the dense neighborhood feature of a network. In small-world networks, the average clustering coefficient is much higher than the random networks.[51,52] Characteristic path length and the average characteristic path length reflect the compactness of a network. In small-world networks, the average characteristic path length is smaller than or comparable to the random networks.[52]

The topological properties (power law degree distribution, clustering coefficient and characteristic path length) of the mirSRNs and miRNA-miRNA synergistic networks are analyzed using NetworkAnalyzer,[53] one of the Cytoscape[54] plugins for computing topological parameters of biological networks. In the analysis, edges of mirSRNs and miRNA-miRNA synergistic networks are treated as undirected because the degree of a node is defined as the number of edges connected with the node, without considering the directions of the edges. Node degree distribution, $p(k)$ is defined as the number of nodes with the node degree k (k=0, 1, 2,…). The dependency between a node degree and the number of nodes with the degree can be visualized by fitting a line on node degree distribution data. NetworkAnalyzer considers only data points with positive values for fitting the power law curve of the form $y=bx^a$. The R^2 value is a statistical measure of the linearity of the curve fit and used to quantify the fit to the power line. The maximal value of R^2 is 1. The larger the R^2 value is, the better the fit is. If the node degree distribution of a network follows a power law, asymptotically, it is a scale-free network.

In addition, as for mirSRNs and miRNA-miRNA synergistic networks (and random networks), the topological measurements (the average clustering coefficient and the average characteristic path length) are obtained using the RandomNetworks plugin (version 0.1) of Cytoscape. To determine if mirSRNs and miRNA-miRNA synergistic networks are small-world networks, we use the duplication model,[55] a well-known model having power-law degree distributions and providing small-world networks to generate random networks. We construct 10000 instances and compute the mean shortest path length and average clustering coefficient.

In order to discover miRNA clusters (highly interconnected regions) in miRNA-miRNA synergistic networks, the MCODE software[56] is used. Instead of directly using clustering coefficient, MCODE uses core-clustering coefficient which measures the density of the highest *k*-core in the neighborhood of a vertex. This amplifies the weighting of densely connected regions in a biological network.

We infer *hub* miRNAs in the miRNA-miRNA synergistic networks, since highly connected nodes (*hubs*) play important roles in different biological networks. Considering that *hubs* may be varied in terms of network size, we don't directly define *hubs* with more than a given node degree, e.g., 10. Similar to the method of Jiang *et al.*,[57] we use the following formula to determine whether a miRNA is a *hub* miRNA:

$$p(x \ge k) = 1 - p(x < k) = 1 - \sum_{i=0}^{k-1} \frac{e^{-\lambda}\lambda^{i}}{i!} \qquad (1)$$

where $\lambda = np_1$, $p_1 = m / A_n^2$, n is the number of miRNAs, m is the number of miRNA causal interaction pairs in the miRNA-miRNA synergistic network, and A_n^2 is the number of all possible miRNA-miRNA synergistic pairs. The smaller the calculated *p*-value of a miRNA is, the more likely the miRNA is a *hub* miRNA. We regard a miRNA with *p*-value < 0.05 as a *hub* miRNA.

2.4 Functional analysis of synergistic miRNAs

To determine whether the synergistic miRNAs are disease miRNAs or not, we conduct validation using the Human MicroRNA Disease Database (HMDD v2.0, http://cmbi.bjmu.edu.cn/hmdd), a curated database for human miRNA and disease associations.[58] To further understand the biological functions and diseases associated with synergistic miRNAs, we use the TAM[59] software to make a functional analysis of them. Significant biological functions and associated diseases are identified for synergistic miRNAs with an adjusted *p*-value (adjusted by Bonferroni method) cutoff of 0.05.

3 Results

3.1 Implementation

In this work, the *limma* package[60] from Bioconductor is used to make the differentially expressed gene analysis. We also remove those genes with more than 10 missing values or identical values. As a result of the analysis, we identify 46 probes of miRNAs, 112 probes of TFs and 1500 probes of mRNAs for the EMT dataset, and 66 probes of miRNAs, 104 probes of TFs and 1214 probes of mRNAs are found for the MCC dataset, at the significant level (adjusted *p*-value<0.05, adjusted by Benjamini-Hochberg (BH) method). The detailed results can be found in the Supplemental Material 1 (ESI†).

The layout of an input dataset or matrix for our framework has two types. As for discovering TF-miRNA causal relationships, the layout of the input dataset/matrix is that the first set of columns are TF expression data, the next set of columns are miRNA expression data and the last set of columns are mRNA expression data. However, as for finding miRNA-miRNA causal effects, and miRNA-TF, miRNA-mRNA causal relationships, the layout of the input dataset/matrix is that the first set of columns are miRNA expression data, the next set of columns are TF expression data and the last set of columns are mRNA expression data. The PC algorithm[44] is used to learn causal structure using an input dataset. We use the implementation of the PC algorithm of the open source R-package, *pcalg*[61] and set the significant level of the conditional independence test *α*=0.01. Since a small number of samples can cause unstable estimation of causal effects, we use the bootstrapping method to estimate casual effects of the discovered causal regulations. The number of bootstrapping *M* is set 100 and the median of 100 estimates for each regulation is used as the final result.

3.2 mirSRNs and their topological properties

The mirSRNs are constructed following the workflow shown in Fig.2. The mirSRN of the EMT data set contains 3882 significant causal relationships between 45 miRNAs and 555 unique molecular nodes (TFs and mRNAs). The mirSRN of the MCC data set contains 59 miRNAs and 846 unique molecular nodes (TFs and mRNAs), and 7555 significant causal relationships. The detailed information of the mirSRNs of EMT and MCC can be seen in Supplemental Material 2, ESI†.

In the mirSRN of EMT, the top 10 miRNAs with the largest node degrees are miR-32, let-7g, miR-7, miR-200c, miR-141, miR-203, miR-96, miR-429, miR-590-3p and miR-101. Among them, 3 miRNAs (miR-200c, miR-141 and miR-429) are validated to play a critical role in the suppression of EMT.[62-64] To further understand whether the top 10 miRNAs are associated with diseases of the EMT dataset or not, we use the HMDD database (Version v2.0) for validation. We find that the top 10 miRNAs, except miR-590-3p, are all closely associated with different human diseases of the EMT dataset (details in Supplemental Material 3, ESI†).

Table 1 Network parameters of mirSRNs of the EMT and MCC datasets.

mirSRN	Nodes	Edges	Clustering coefficient	Characteristic path length	$y=bx^a$	R^2
EMT	600	3882	0.804	2.389	y=81.339$x^{-0.957}$	0.702
MCC	905	7555	0.735	2.386	y=149.28$x^{-1.008}$	0.742

In the mirSRN of MCC, miR-195, miR-26a, let-7g, miR-30a, miR-107, miR-130a, miR-103, miR-126, miR-24, and miR-30c are the top 10 miRNAs with the largest node degrees. The validation results by HMDD demonstrate that the top 10 miRNAs except miR-103, are all associated with different human diseases of the MCC dataset (details in Supplemental Material 3, ESI†).

We then analyze the network properties of the mirSRNs of the EMT and MCC datasets by NetworkAnalyzer. The distributions of node degrees of the two networks approximately follow power law distributions, with R^2 = 0.702 and 0.742, respectively (Table 1). Therefore, the mirSRNs are approximately scale-free, which is one of most important characteristics of true complex biological networks.[50]

For the mirSRNs of the EMT and MCC datasets, the average clustering coefficients are 0.804 and 0.735 respectively, which are much higher than that for a random network (0.040 ± 0.010 and 0.030 ± 0.007). Moreover, the average characteristic path lengths of the two mirSRNs are 2.389 and 2.386 respectively, which are lower than those of random networks generated by the duplication model (3.852 ± 0.077 and 4.027 ± 0.074). This result indicates that the mirSRNs of the EMT and MCC datasets are small-world networks with high clustering coefficients and small characteristic path lengths,[51,52] and the synergistic miRNAs can promptly implement gene regulation.

In summary, the top 10 miRNAs with largest node degrees in each mirSRN are almost all associated with different human diseases of the EMT or MCC datasets, including cancers, and the network is approximately scale-free and small-world.

3.3 Evaluation of miRNA-miRNA synergistic networks

The miRNA-miRNA synergistic network of the EMT dataset (Fig. 3(A)) contains 45 miRNAs and 1012 edges, while that of the MCC dataset (Fig. 3(B)) contains 59 miRNAs and 1699 edges. We divide the miRNAs into Disease miRNAs and Non-disease miRNAs following the HMDD database. In the EMT network, 10 out of 45 miRNAs are Non-disease miRNAs (miR-1307, miR-17*, miR-192*, miR-200a*, miR-200c*, miR-331-3p, miR-380, miR-590-3p, miR-590-5p and miR-769-5p). All the miRNAs of the MCC network except miR-103 are all Disease miRNAs.

The results imply that the EMT and MCC miRNA-miRNA synergistic networks are also small-world. From Fig. 3 (tables at the bottom of the figure), the distributions of node degrees of the EMT and MCC miRNA-miRNA synergistic networks do not follow power law distributions, with R^2 = 0.028 and 0.019, respectively. Therefore, the EMT and MCC miRNA-miRNA synergistic networks are not scale-free. The topologies of the two networks exhibit dense local neighbourhoods with the average clustering coefficients of 0.789 and 0.770 respectively, which are much higher than those of random networks (0.190 ± 0.067 and 0.167 ± 0.056). We also find that most miRNAs are connected together and the two networks have small average characteristic path lengths of 1.360 and 1.298, respectively. These values are lower than those of random graphs generated by the duplication model (2.657 ± 0.096 and 2.787 ± 0.096). Since the dense neighbourhood feature and small characteristic path length can be exploited to predict synergism,[51,52]

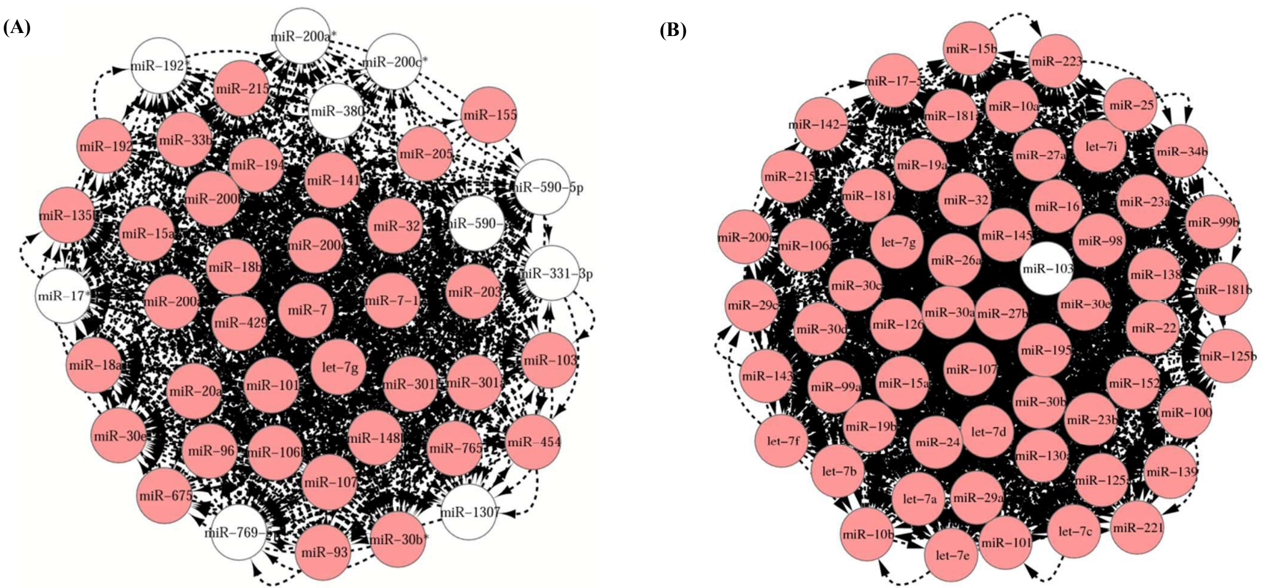
The result implies that most miRNAs are frequently synergistic with their miRNA partners in the miRNA-miRNA synergistic networks. We evaluate the *hub* miRNAs of the EMT and MCC miRNA-miRNA synergistic networks. The percentages of *hub* miRNAs in the two networks are 82.22% and 81.36% respectively, and the detailed results of node degree and *p*-value for each miRNA can be seen in Supplemental Material 4 (ESI†).

In sum, all miRNAs are closely connected in each network, the networks are not scale-free but small-world and most miRNAs are frequently synergistic with their miRNA partners.

3.4 miRNAs associated with the same disease are likely to be synergistic and form clusters

In this section, we focus on studying the relationships between miRNA clusters and diseases. The MCODE software is applied to find clusters (highly interconnected regions) in the miRNA-miRNA synergistic networks, which would provide important insight to the miRNAs involved. As shown in Table 2, we identify 2 clusters (Score>1) in both the EMT and MCC miRNA-miRNA synergistic networks. For the EMT dataset, all the 47 samples are closely related to 9 human cancer cell lines (Breast, Cardiovascular Nervous System, Colon, Leukemia, Lung, Melanoma, Ovarian, Prostate and Renal). Meanwhile, all the 88 samples in the MCC dataset are closely associated with 11 human cancer lines (Bladder, Breast, Colon, Lung, Melanoma, Mesothelioma, Ovarian, Pancreas, Prostate, Renal and Uterus). Therefore, we are only interested in miRNA clusters associated with the above 9 and 11 human diseases in the EMT and MCC datasets, respectively.

Two clusters containing 41 unique miRNA are found in the EMT miRNA-miRNA synergistic network and two clusters containing 53 unique miRNAs are found in the MCC network. From the HMDD database (Version v2.0), the 2 clusters in the EMT network are involved in the 9 human diseases in the EMT dataset, and the 2 clusters in the MCC network are associated with the 11 human diseases in the MCC dataset.



Nodes	45	Edges	1012
Characteristic path length	1.360	Clustering coefficient	0.789
y=*bx*^{a}	*y*=0.880*x*$^{0.109}$	R^2	0.028

Nodes	59	Edges	1699
Characteristic path length	1.298	Clustering coefficient	0.770
y=*bx*^{a}	*y*=3.412*x*$^{-0.197}$	R^2	0.019

Fig. 3 Graphic representations of miRNA-miRNA synergistic networks (generated using Cytoscape). (A) The miRNA-miRNA synergistic network of the EMT dataset. (B) The miRNA-miRNA synergistic network of the MCC dataset. The Disease miRNA nodes are colored in red, and the Non-disease miRNA nodes are colored in white.

For the EMT network, more than half of the 31 miRNAs in Cluster 1 are associated with Breast Neoplasms (20 miRNAs), Colonic Neoplasms (19 miRNAs), Lung Neoplasms (17 miRNAs) and Melanoma (17 miRNAs). Meanwhile, half of the 10 miRNAs in Cluster 2 are related to Colonic Neoplasms.

For the MCC network, more than half of the total 40 miRNAs in Cluster 1 are related to Breast Neoplasms (31 miRNAs), Colonic Neoplasms (32 miRNAs), Lung Neoplasms (32 miRNAs), Melanoma (28 miRNAs), Ovarian Neoplasms (26 miRNAs), Pancreatic Neoplasms (22 miRNAs), Prostate Neoplasms (26 miRNAs) and Renal (22 miRNAs), respectively. Moreover, more than half of the 13 miRNAs in Cluster 2 are associated with Breast Neoplasms (10 miRNAs), Colonic Neoplasms (9 miRNAs), Lung Neoplasms (8 miRNAs), Melanoma (8 miRNAs), Ovarian Neoplasms (8 miRNAs), Pancreatic Neoplasms (7 miRNAs) and Prostatic Neoplasms (7 miRNAs), respectively. The detailed information can be seen in Supplemental Material 5 (ESI†).

In all, these results indicate that miRNAs associated with the same disease are likely to be synergistic and form clusters to implement biological functions.

3.5 Functional validation of synergistic miRNAs

To understand the biological functions and diseases associated with synergistic miRNAs, we use the TAM[59] software to make a functional analysis of the synergistic miRNAs contained in the miRNA-miRNA synergistic networks of the EMT and MCC datasets. Similar to the cluster analysis of the miRNA-miRNA synergistic networks, for the EMT dataset, we are only interested in the links between the synergistic miRNAs and the epithelial-mesenchymal transition, and diseases associated with the 9 human cancer cell lines in the dataset; and for the MCC dataset, we are only concerned about miRNA tumor suppressors in biological function, and the diseases with associated the 11 human cancer cell lines in the dataset.

Table 2 Cluster analysis of miRNA-miRNA synergistic networks of the EMT and MCC datasets.

Dataset	Cluster	Score (Density*#Nodes)	Nodes	Edges	miRNAs
EMT	1	21.806	31	676	miR-30e, miR-194, miR-429, miR-200c, miR-203, miR-141, miR-200b, miR-17*, miR-200a, miR-107, miR-32, miR-18b, miR-301a, miR-7, miR-106b, miR-590-3p, miR-148b, miR-18a, miR-101, miR-331-3p, miR-301b, miR-135b, miR-96, miR-7-1*, miR-33b, let-7g, miR-30b*, miR-765, miR-590-5p, miR-15a, miR-769-5p
	2	2.9	10	29	miR-215, miR-192*, miR-200a*, miR-192, miR-675, miR-93, miR-103, miR-454, miR-1307, miR-20a
MCC	1	25.35	40	1014	miR-30e, miR-17-5p, miR-99b, miR-19a, miR-223, miR-34b, miR-19b, miR-15a, miR-24, miR-29c, miR-30d, miR-181a, miR-107, miR-106a, miR-138, let-7e, miR-181c, miR-145, miR-27b, miR-29a, miR-126, let-7g, miR-26a, miR-152, miR-30c, let-7f, miR-30b, miR-32, miR-195, miR-101, miR-99a, miR-16, miR-23b, let-7c, miR-98, miR-30a, miR-143, miR-130a, let-7i, miR-27a
	2	3.846	13	50	miR-103, miR-139, miR-25, miR-23a, miR-142-3p, let-7b, miR-100, miR-125a, miR-22, miR-215, miR-15b, let-7d, let-7a

As a result of the analysis, 15 out of the 45 miRNAs in the EMT miRNA-miRNA synergistic network are associated with epithelial-mesenchymal transition. The numbers of the miRNAs closely related to Breast Neoplasms, Colonic Neoplasms, Lung Neoplasms, Melanoma, Ovarian Neoplasms and Prostatic Neoplasms are 22, 13, 18, 19, 17 and 15 in EMT, respectively (see Supplemental Material 6, ESI†).

As for the MCC miRNA-miRNA synergistic network, 26 out of the 59 miRNAs in the network are related to miRNA tumor suppressors, and 29, 24, 38, 31, 31, 25, 22, 12 and 9 miRNAs are associated with Breast Neoplasms, Colonic Neoplasms, Lung Neoplasms, Melanoma, Ovarian Neoplasms, Pancreatic Neoplasms, Prostatic Neoplasms, Carcinoma of Renal Cell and Uterine Cervical Neoplasms, respectively (see Supplemental Material 6, ESI†). These results demonstrate that the identified synergistic miRNAs can cause different kinds of diseases.

3.6 Synergistic miRNA pairs have higher co-expression level

Synergistic miRNA pairs tend to be co-expressed. To validate our assumption, we investigate the co-expression levels of synergistic miRNA pairs in the miRNA-miRNA synergistic networks. We use Pearson correlation of each synergistic miRNA pair to measure the (relevant) level of their co-expression. If a synergistic miRNA pair is positively correlated with p-value < 0.05, the synergistic miRNA pair is regarded as a co-expressed miRNA pair; otherwise the pair is not co-expressed. For comparison, we also calculate Pearson correlation of each non-synergistic miRNA pair to know their co-expression level. To further investigate whether the synergistic miRNA pairs have higher levels of co-expression or not, we use the Kolmogorov-Smirnov (KS) test to evaluate the difference between the co-expression levels of synergistic miRNA pairs and the co-expression levels of non-synergistic miRNA pairs in each of the EMT and MCC miRNA-miRNA synergistic networks. Please note that, when evaluating the difference, we only consider co-expressed miRNA pairs of the synergistic and non-synergistic miRNA pairs.

As illustrated in Fig. 4, the synergistic miRNA pairs have higher co-expression levels than those of non-synergistic miRNA pairs in the EMT and MCC miRNA-miRNA synergistic networks with p-value = 9.9939E-110 and 8.6809E-058, respectively. The detailed results of co-expression miRNA pairs of the synergistic and non-synergistic miRNA pairs of each network are provided in Supplemental Material 7 (ESI†).

The above results support the proposition that the synergistic miRNA pairs have higher levels of co-expression, and they have potential functional relationships which may indicate collaboration between the miRNAs. The higher level of co-expression may demonstrate that miRNAs can quickly adapt to a new biological environment through prompt gene regulation.

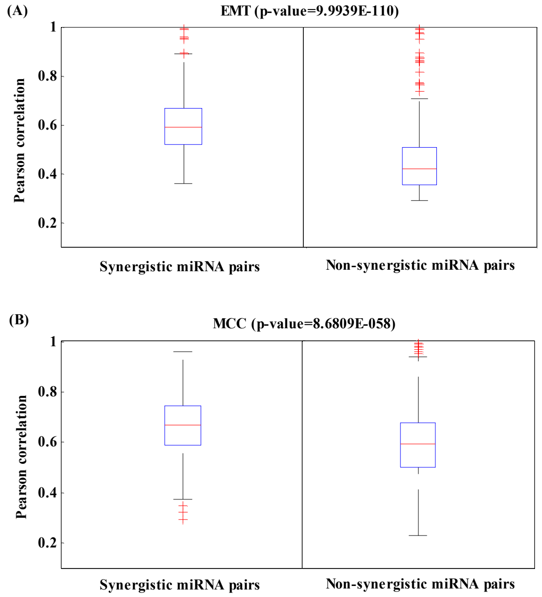


Fig. 4 Co-expression level of synergistic and non-synergistic miRNA pairs in the miRNA-miRNA synergistic networks. (A) Co-expression level of synergistic and non-synergistic miRNA pairs in the EMT network. (B) Co-expression level of synergistic and non-synergistic miRNA pairs in the MCC network. In both EMT and MCC networks, the synergistic miRNA pairs have higher co-expression level than that of non-synergistic miRNA pairs. The p-values are calculated using the Kolmogorov-Smirnov (KS) test.

4 Conclusions

miRNAs are main regulators at the post-transcriptional level, and they play important roles in most biological processes, including cancers. Previous research[12-14] has demonstrated that multiply miRNAs may work synergistically to regulate their target genes. It is important to study the synergism of miRNAs at a system-wide level to further understand the regulation mechanism of miRNAs.

In this study, we propose a framework called mirSRN to construct the miRNA synergistic regulatory networks from heterogeneous human data, including sequence and expression data, and our framework uses a causal discovery model to infer mirSRNs.

We hypothesize that miRNAs can both directly and indirectly cooperate with each other to regulate their targets. Consequently, we have two types of mirSRN motifs: DmirSRN motifs and ImirSRN motifs. Significant causal relationships of miRNA-mRNA and miRNA-TF are used to generate DmirSRN motifs, while the significant TF-miRNA and miRNA-TF causal relationships are used to develop ImirSRN motifs. The strength of miRNA synergism is measured by the miRNA-miRNA causal effect.

The mirSRNs of the EMT and MCC datasets show the desirable generic properties of general biological networks. The two networks are approximately scale-free and small-world. The top 10 miRNAs with largest node degrees are almost all associated with different human diseases of the EMT and MCC datasets, including cancers.

To further infer the properties of synergistic miRNAs, we construct the miRNA-miRNA synergistic networks by extracting only the miRNA-miRNA synergistic relationships from the mirSRNs. We find that the miRNA-miRNA synergistic networks have the following features: all miRNAs are closely connected in each network, the networks are not scale-free but small-world and most miRNAs are frequently synergistic with their miRNA partners.

Moreover, we make cluster analysis using the MCODE software and find that miRNAs associated with the same disease are likely to be synergistic and link the same biological functions.

The comparison results between the co-expression levels of synergistic and non-synergistic miRNA pairs demonstrate that synergistic miRNA pairs show higher co-expression level, and may have potential functional relationships. According to the functional analysis of synergistic miRNAs by the TAM software, we find that the identified synergistic miRNAs link different kinds of diseases.

In conclusion, the results from the proposed framework provide new insights into understanding miRNA synergistic regulation mechanisms, especially disease miRNAs. The framework has great potential to improve our understanding of the roles of miRNAs in different kinds of diseases.

Author's contribution

JPZ, TDL and JYL conceived the idea of this work. LL and JFH refined the idea. JPZ and TDL designed and performed the experiments. JPZ, TDL, LL and JYL drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the Applied Basic Research Foundation of Science and Technology of Yunnan Province (No: 2013FD038), and the Australian Research Council Discovery grant (DP130104090).

References

1 K. Chen and N. Rajewsky, *Nat. Rev. Genet.*, 2007, **8**, 93–103.

2 D. P. Bartel, *Cell*, 2009, **136**, 215–33.

3 B. P. Lewis, C. B. Burge and D. P. Bartel, *Cell*, 2005, **120**, 15-20.

4 R. C. Friedman, K. K. Farh, C. B. Burge and D. P. Bartel, *Genome Res.*, 2009, **19**, 92-105.

5 V. Ambros, *Cell*, 2001, **107**, 823-6.

6 V. Ambros, *Cell*, 2003, **113**, 673-6.

7 D. P. Bartel, *Cell*, 2004, **116**, 281-97.

8 N. Bushati and S.M. Cohen, *Annu. Rev. Cell Dev. Biol.*, 2007, **23**, 175-205.

9 T. Du and P. D. Zamore, *Cell Res.*, 2007, **17**, 661-3.

10 A. Esquela-Kerscher and F. J. Slack, *Nat. Rev. Cancer*, 2006, **6**, 259-269.

11 Q. Cui, Z. Yu, E. O. Purisima and E. Wang, *Mol. Syst. Biol.*, 2006, **2**, 46.

12 A. J. Enright, B. John, U. Gaul, T. Tuschl, C. Sander and D. S. Marks, *Genome Biol.*, 2003, **5**, R1.

13 A. Krek, D. Grün, M. N. Poy, R. Wolf, L. Rosenberg, E. J. Epstein, P. MacMenamin, I. da Piedade, K. C. Gunsalus, M. Stoffel and N. Rajewsky, *Nat. Genet.*, 2005, **37**, 495-500.

14 S. Wu, S. Huang, J. Ding, Y. Zhao, L. Liang, T. Liu, R. Zhan and X. He, *Oncogene*, 2010, **29**, 2302-8.

15 G. Sales, A. Coppe, A. Bisognin, M. Biasiolo, S. Bortoluzzi and C. Romualdi, *Nucleic Acids Res.*, 2010, **38**, W352-9.

16 J. B. Hsu, C. M. Chiu, S. D. Hsu, W. Y. Huang, C. H. Chien, T. Y. Lee and H. D. Huang, *BMC Bioinformatics*, 2011, **12**, 300.

17 A. Ferro, R. Giugno, A. Laganà, M. Mongioví, G. Pigola, A. Pulvirenti, G. Bader and D. Shasha, in *Network Tools and Applications in Biology (NETTAB), Focused on Technologies, Tools and Applications for Collaborative and Social Bioinformatics Research and Development*, ed. C. Romano, Catania: Libero di Scrivere, 2009.

18 M. Kutmon, T. Kelder, P. Mandaviya, C. T. Evelo and S. L. Coort, *PLoS One*, 2013, **8**, e82160.

19 G. Politano, A. Benso, A. Savino and S. Di Carlo, *PLoS One*, 2014, **9**, e115585.

20 X. Yuan, C. Liu, P. Yang, S. He, Q. Liao, S. Kang and Y. Zhao, *BMC Syst. Biol.*, 2009, **3**, 65.

21 J. An, K. P. Choi, C. A. Wells and Y. P. Chen, *J. Bioinform. Comput. Biol.*, 2010, **8**, 99-115.

22 J. Xu, C. X. Li, Y. S. Li, J. Y. Lv, Y. Ma, T. T. Shao, L. D. Xu, Y. Y. Wang, L. Du, Y. P. Zhang, W. Jiang, C. Q. Li, Y. Xiao and X. Li, *Nucleic Acids Res.*, 2011. **39**, 825-836.

23 D. Sengupta and S. Bandyopadhyay, *Mol. Biosyst.*, 2011, **7**, 1966-73.

24 W. Zhu, Y. Zhao, Y. Xu, Y. Sun, Z. Wang, W. Yuan and Z. Du, *PLoS One*, 2013, **8**, e63342.

25 J. Xu, Y. Li, X. Li, C. Li, T. Shao, J. Bai, H. Chen and X. Li, *Mol. Biosyst.*, 2013, **9**, 217-24.

26 X. Kong, X. Geng, J. Zhao and Y. Hu, *Int. J. Biosci. Biochem. Bioinforma.*, 2015, **5**, 184.

27 G. Boross, K. Orosz and I. J. Farkas, *Bioinformatics*, 2009, **25**, 1063-1069.

28 M. Alshalalfa, *Adv. Bioinformatics*, 2012, **2012**, 839837.

29 X. Zhao, H. Song, Z. Zuo, Y. Zhu, X. Dong and X. Lu, *Int. J. Biol. Macromol.*, 2013, **55**, 98-103.

30 L. Hua, H. Xia, P. Zhou, D. Li and L. Li, *Biosci. Trends*, 2014, **8**, 297-307.

31 L. Guo, Y. Zhao, S. Yang, H. Zhang and F. Chen, *Biomed. Res. Int.*, 2014, **2014**, 907420.

32 Y. Li, C. Liang, K. C. Wong, J. Luo and Z. Zhang, *Bioinformatics*, 2014, **30**, 2627-35.

33 K. Dimitrakopoulou, A. G. Vrahatis, A. Bezerianos, *BMC Genomics*, 2015, **16**, 147.

34 R. Søkilde, B. Kaczkowski, A. Podolska, S. Cirera, J. Gorodkin, S. Møller and T. Litman, *Mol. Cancer Ther.*, 2011, **10**, 375-84.

35 J. Lu, G. Getz, E. A. Miska, E. Alvarez-Saavedra, J.Lamb, D.Peck, A. Sweet-Cordero, B. L. Ebert, R. H. Mak, A. A. Ferrando, J. R. Downing, T. Jacks, H. R. Horvitz and T. R. Golub, *Nature*, 2005, **435**, 834-8.

36 S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander and T. R. Golub, *Proc. Natl. Acad. Sci. USA*. 2001, **98**, 15149-54.

37 J. M. Vaquerizas, S. K. Kummerfeld, S. A. Teichmann and N. M. Luscombe, *Nat. Rev. Genet.*, 2009, **10**, 252-63.

38 S. Griffiths-Jones, H. K. Saini, S. van Dongen and A. J. Enright, *Nucleic Acids Res*., 2008, **36**, D154-8.

39 V. Agarwal, G. W. Bell, J. W. Nam and D. P. Bartel, *Elife*, 2015, **4**, e05005.

40 J. H. Li, S. Liu, H. Zhou, L. H. Qu and J. H. Yang, *Nucleic Acids Res.*, 2014, **42**, D92-7.

41 A. Le Béchec, E. Portales-Casamar, G. Vetter, M. Moes, P. J. Zindy, A. Saumet, D. Arenillas, C. Theillet, W. W. Wasserman, C. H. Lecellier and E. Friederich, *BMC Bioinformatics*, 2011, **12**, 67.

42 H. M. Maathuis, M. Kalisch and P. Buhlmann, *Ann. Stat.*, 2009, **37**, 3133-3164.

43 H. M. Maathuis, D. Colombo, M. Kalisch and P. Buhlmann, *Nat. Methods*, 2010, **7**, 247-249.

44 P. Spirtes, C. Glymour and R. Scheines, *Causation, Prediction, and Search*, MIT Press, Cambridge, 2000.

45 P. Judea, *Causality: Models, Reasoning, and Inference*, Cambridge University Press, Cambridge, 2000.

46 T. D. Le, L. Liu, A. Tsykin, G. J. Goodall, B. Liu, B. Y. Sun and J. Li, *Bioinformatics*, 2013, **29**, 765-771.

47 R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon, *Science*, 2002, **298**, 824–827.

48 J. F. Knabe, C. L. Nehaniv and M. J. Schilstra, *Biosystems*, 2008, **94**, 68-74.

49 A. Ferro, R. Giugno, G. Pigola, A. Pulvirenti, D. Skripin, G. D. Bader and D. Shasha, *Bioinformatics*, 2007, **23**, 910-2.

50 A. L. Barabási and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, 101-13.

51 D. S. Goldberg and F. P. Roth, *Proc. Natl. Acad. Sci. U S A*, 2003, **100**, 4372-6.

52 Q. K. Telesford, K. E. Joyce, S. Hayasaka, J. H. Burdette and P. J. Laurienti, *Brain Connect.*, 2011, **1**, 367-75.

53 Y. Assenov, F. Ramírez, S. E. Schelhorn, T. Lengauer and M. Albrecht, *Bioinformatics*, 2008, **24**, 282-4.

54 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498-504.

55 A. L. Barabasi and R. Albert, *Science*, 1999, **286**, 509-12.

56 G. D. Bader and C. W. Hogue, *BMC Bioinformatics*, 2003, **4**, 2.

57 W. Jiang, X. Li, S. Rao, L. Wang, L. Du, C. Li, C. Wu, H. Wang, Y. Wang and B. Yang, *BMC Syst. Biol.*, 2008, **2**, 72.

58 M. Lu, Q. Zhang, M. Deng, J. Miao, Y. Guo, W. Gao and Q. Cui, *PLoS One*, 2008, **3**, e3420.

59 M. Lu, B. Shi, J. Wang, Q. Cao and Q. Cui, *BMC Bioinformatics*, 2010, **11**, 419.

60 M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, and G. K. Smyth, (2015). *Nucleic Acids Res.*, 2015, **43**, e47.

61 M. Kalisch, M. Machler, D. Colombo, M. H. Maathuis and P. Bühlmann, *J. Stat. Softw.*, 2012, **47**, 1-26.

62 P. A. Gregory, C. P. Bracken, A. G. Bert and G. J. Goodall, *Cell Cycle*, 2008, **7**, 3112-8.

63 M. Korpal and Y. Kang, *RNA Biol.*, 2008, **5**, 115-9.

64 P. S. Mongroo and A. K. Rustgi, *Cancer Biol. Ther.*, 2010, **10**, 219-22.