Journal of Biomedical Informatics 52 (2014) 438–447

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Identifying direct miRNA–mRNA causal regulatory relationships in heterogeneous data

Junpeng Zhang [a,1], Thuc Duy Le [b,1], Lin Liu [b], Bing Liu [c], Jianfeng He [d], Gregory J. Goodall [e], Jiuyong Li [b,*]

[a] Faculty of Engineering, Dali University, Dali, China
[b] School of Information Technology and Mathematical Sciences, University of South Australia, Mawson Lakes, SA 5095, Australia
[c] Children's Cancer Institute Australia, Randwick, NSW 2301, Australia
[d] Kunming University of Science and Technology, Kunming, China
[e] Centre for Cancer Biology, SA Pathology, SA 5000, Australia

ARTICLE INFO

Article history:
Received 26 December 2013
Accepted 16 August 2014
Available online 30 August 2014

Keywords:
MicroRNA
mRNA
Causal regulatory relationships
Epithelial–mesenchymal transition
Causal feedforward patterns

ABSTRACT

Discovering the regulatory relationships between microRNAs (miRNAs) and mRNAs is an important problem that interests many biologists and medical researchers. A number of computational methods have been proposed to infer miRNA–mRNA regulatory relationships, and are mostly based on the statistical associations between miRNAs and mRNAs discovered in observational data. The miRNA–mRNA regulatory relationships identified by these methods can be both direct and indirect regulations. However, differentiating direct regulatory relationships from indirect ones is important for biologists in experimental designs. In this paper, we present a causal discovery based framework (called DirectTarget) to infer direct miRNA–mRNA causal regulatory relationships in heterogeneous data, including expression profiles of miRNAs and mRNAs, and miRNA target information. DirectTarget is applied to the Epithelial to Mesenchymal Transition (EMT) datasets. The validation by experimentally confirmed target databases suggests that the proposed method can effectively identify direct miRNA–mRNA regulatory relationships. To explore the upstream regulators of miRNA regulation, we further identify the causal feedforward patterns (CFFPs) of TF–miRNA–mRNA to provide insights into the miRNA regulation in EMT. DirectTarget has the potential to be applied to other datasets to elucidate the direct miRNA–mRNA causal regulatory relationships and to explore the regulatory patterns.

1. Background

A fundamental challenge of understanding complex gene regulatory mechanisms is to identify how the regulators regulate their target genes in different biological processes, especially in disease progression. Many studies have demonstrated that microRNAs (miRNAs) are primary metazoan gene regulators at the post-transcriptional level. miRNAs are short (~22 nt) endogenous non-coding RNAs and recognise target genes by binding to complementary sequences on the target messenger RNA (mRNAs) transcripts. The binding activities usually result in translational repression or target degradation and gene silencing [1,2]. The research into miRNAs has revealed the roles that they play in negative regulation and possibly in positive regulation. By regulating target genes, miRNAs are likely to be involved in most biological processes, including developmental timing, cell proliferation, metabolism, differentiation, apoptosis, cellular signaling, stress responses and cancers [3–9].

miRNA target prediction using sequence data [10–13] can provide the direct reference of miRNA binding sites on the target genes. However, these methods are based on sequence complementarity and/or structural stability of the putative duplex. This may result in a high rate of false positives and false negatives [14], mainly because of the role played by RNA folding as well as accessibility due to protein binding. Moreover, the currently available sequence-based target prediction algorithms produce hundreds to a few thousands of target genes for each miRNA, which makes it difficult to focus on a small number of most likely targets of the miRNAs of interest.

Complementary to the target prediction based on sequence data, some methods have been developed to use gene expression data or the combination of expression data and sequence data to find out miRNA targets. For example, the approaches presented

* Corresponding author. Tel.: +61 8 83023898; fax: +61 8 83023381.
E-mail addresses: zhangjunpeng_411@yahoo.com (J. Zhang), leyty017@mymail.unisa.edu.au (T.D. Le), lin.liu@unisa.edu.au (L. Liu), BLiu@ccia.unsw.edu.au (B. Liu), jfenghe@kmust.edu.cn (J. He), greg.goodall@health.sa.gov.au (G.J. Goodall), jiuyong.li@unisa.edu.au (J. Li).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.jbi.2014.08.005
1532-0464/

in [15–18] identify miRNA regulatory modules (MRMs) by integrating heterogeneous datasets, including expression profiles of miRNAs and mRNAs, and putative target binding information. The discovery of MRMs has demonstrated that using both sequence information and expression profiles can produce more accurate predictions. Attempts have also been made to infer functional miRNA–mRNA regulatory modules (FMRMs), which are regulatory networks of miRNAs and mRNAs for specific biological conditions [19–26]. The identified FMRMs give insights into biological processes, functional regulatory interactions of many diseases and gene target therapy. The limitation shared by these methods is that they can only infer statistical correlations or associations between miRNAs and mRNAs. However, correlations or associations may not reveal gene regulatory relationships which are indeed causal relationships. For example, a strong correlation between miRNA *A* and a target gene *B* does not necessarily imply that *A* causally regulates *B*. This strong correlation between *A* and *B* may be due to a common cause (regulator) of them.

Recently Le et al. [27] adapted the causal modelling and discovery approach, IDA [28,29] to infer miRNA–mRNA *causal* regulatory relationships from gene expression data. The discovered miRNA–mRNA causal regulatory relationships were found to have a large portion of overlap with the results of the follow-up gene knockdown experiments. This outcome has demonstrated the high accuracy that can be achieved by the computational method. However, the method does not distinguish between direct and indirect causal relationships. Therefore, a discovered causal regulatory relationship can be a direct interaction between a miRNA and a mRNA, or indirect regulation of a miRNA on a mRNA that is mediated by some other regulator(s). Given that a major benefit of identifying miRNA–mRNA causal regulatory relationships is to provide biologists with high quality hypothetical miRNA target information to assist them in setting up gene knockdown experiments, it is essential to have direct miRNA–mRNA causal regulatory relationships identified and presented to biologists.

In this paper, we extend the work in [27] to infer *direct* miRNA–mRNA causal regulatory relationships. Our framework (called DirectTarget) makes use of multiple sources of data, including gene expression profiles and target binding information. We hypothesise that target information predicted using sequence data provides evidence of direct interactions between miRNAs and the predicted targets. With DirectTarget, firstly we apply the approach in [27] to the expression profiles of miRNAs and mRNAs to identify all the pairs of miRNA and mRNA that are causally related. Then based on the target information, we select from the pairs only direct causal relationships. We then use experimentally confirmed target databases to validate the computational results.

Since transcription factors (TFs) are main regulators at the transcriptional level, they are also essential for the regulation of gene expression, including miRNA expression levels. Therefore, to further explore the upstream regulators (TFs) of miRNA regulation, we repeat the procedure of DirectTarget for inferring the direct TF–miRNA and TF–mRNA causal regulatory relationships, where miRNAs and mRNAs are targets of the TFs. Then they are integrated with the direct miRNA–mRNA regulatory relationships found to generate the causal feedfoward patterns (CFFPs).

We apply DirectTarget to the EMT datasets and use experimentally confirmed target databases to validate the identified direct causal relationships. Validation results show that the proposed method is suitable for discovering direct miRNA–mRNA causal regulatory relationships. Moreover, the enrichment analyses on the top ranked target genes show that their functions are highly relevant to the biological conditions of the datasets. The computational results of miRNA–mRNA regulatory relationships as well as CFFPs provide good resources for cancer research and future experimental validations.

2. Methods

2.1. Overview

In this section, we present DirectTarget, a framework for inferring direct miRNA–mRNA causal regulatory relationships. As shown in Fig. 1, the overall process contains three steps: (1) data preparation, (2) inferring miRNA–mRNA causal regulatory relationships, and (3) identifying direct miRNA–mRNA causal regulatory relationships. In the following, we present each of the steps in detail.

2.1.1. Step (1): Data preparation

The aim of this step is to obtain the input data for the causal discovery step (Step (2)), including miRNA target binding information and expression profiles of miRNAs and mRNAs. Since we will apply DirectTarget to the Epithelial to Mesenchymal Transition (EMT) datasets, we use the EMT datasets as an example to introduce the data preparation step in the following.

EMT is a process in which cells lose their epithelial features characterised by the high E-cadherin expression level and acquire mesenchymal characteristics, including Vimentin filaments and a flattened phenotype. EMT is a part of the process of tissue remodelling during embryonic development and wound healing [30], and during carcinogenesis [31] when cancer cells undergo a change transforming into a more invasive tumor [30,32]. By expressing proteases, cells become more invasive, and they can pass through the underlying basement membrane and migrate. These are crucial steps in the multi-step process of metastasis [33].

To obtain target binding information of miRNAs, several databases [10–13] may be queried. In this work we use Microcosm (Version v5) [11]. The miRNA expression profiles are from Søkilde et al. [34]. They were 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 for NCI-60 are obtained from ArrayExpress (http://www.ebi.ac.uk/arrayexpress, accessionnumberE-GEOD-5720). In total 47 samples, including 11 epithelial samples and 36 mesenchymal samples are used for this work.

With the retrieved miRNA and mRNA expression profiles, we perform differential gene expression analysis to identify differentially expressed miRNAs and mRNAs between the epithelial and mesenchymal samples. In this work, the *limma* package [35] of Bioconductor is used for the analysis. The expression values of differentially expressed miRNAs and mRNAs are to be used as an input for Step (2).

2.1.2. Step (2): Inferring miRNA–mRNA causal regulatory relationships

This step is based on the causal discovery method, IDA [28,29] and the adaptation of IDA presented in [27] for finding miRNA–mRNA causal regulatory relationships. Therefore two sub-steps need to be carried out: (2a) using the PC algorithm [36] to learn the causal structure from expression data; (2b) applying do-calculus [37] to infer the causal effects of miRNAs on mRNAs.

In Step (2a), each miRNA or mRNA is considered as a random variable whose values are its expression levels. The PC algorithm assumes that there is a true causal structure in the form of a directed acyclic graph (DAG), where a node corresponds to a variable (a miRNA or mRNA in our case) and an edge between two nodes represents the existence of a causal relationship between them. The PC algorithm is used to learn the causal structure (the DAG) from the expression profiles of miRNAs and mRNAs.

The structure learning by the PC algorithm starts with a fully connected undirected graph, assuming that initially every pair of

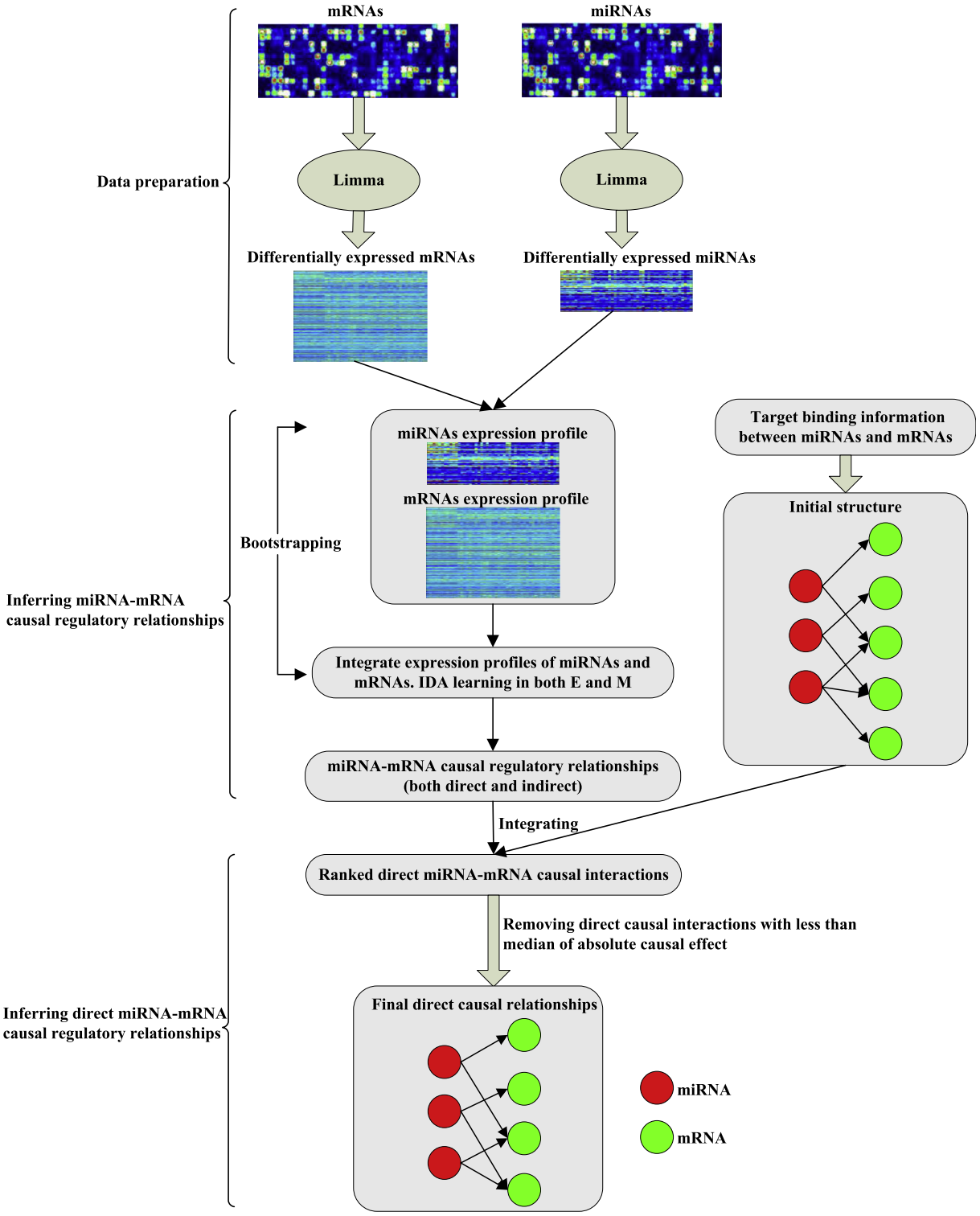

Fig. 1. The workflow (steps) for inferring direct miRNA–mRNA causal regulatory relationships. The overall process of DirectTarget involves three main steps: data preparation, inferring miRNA–mRNA causal regulatory relationships and identifying direct miRNA–mRNA causal regulatory relationships. Putative target binding information between miRNAs and mRNAs, and expression profiles of miRNA and mRNAs are used. Target binding information is used to create the initial structure representing the interactions of miRNA–mRNA. Expression profiles are then used in the IDA learning procedure to construct causal regulatory interactions of miRNA–mRNA in both epithelial (E) and mesenchymal (M) sample conditions. Bootstrapping strategy is used to improve the stability of the estimation of causal effects between miRNAs and mRNAs. The causal interactions and putative target binding information of miRNA–mRNA are further integrated to infer direct miRNA–mRNA causal regulatory relationships. All direct causal interactions are ranked in descending order of the absolute values of the causal effects, and those direct causal interactions with less than median of absolute causal effect are removed when selecting potential direct miRNA–mRNA causal interactions.

nodes (i.e. a pair of miRNA–mRNA, miRNA–miRNA or mRNA–mRNA) is related. Then, the algorithm decides if an edge is to be removed from or retained in the graph by conducting conditional independence tests for the two nodes connected by the edge. Finally, the directions of edges in the obtained graph are oriented with the aim of getting a DAG. However, the PC algorithm can only

generate a *completed partially directed acyclic graph* (CPDAG) which contains both directed edges and undirected edges. From a CPDAG, there are several ways to orient the undirected edges to form a DAG. Therefore, the resulting causal structure is not a unique DAG, but a class of DAGs generated from the CPDAG. Fig. 2 shows an example of a CPDAG, G, and from G there are four different ways to orient the edges, resulting in the four DAGs in the equivalence class [28].

In Step (2b), do-calculus [37] is used to estimate the causal effect that a miRNA has on a mRNA. Given a DAG, do-calculus can estimate the causal effect of a node on any other node in the DAG using observational data, by way of 'simulating' the interventions in controlled experiments. However, the problem is that we do not have just one unique DAG, but a class of DAGs as described above. The solution [27–29] is that to find out the causal effect of a miRNA on a mRNA, we estimate the causal effects using the DAGs resulting from a CPDAG one by one and take the minimum causal effect value as the final result for this pair of miRNA and mRNA. The idea is that, since we are not able to estimate the unique causal effect between each miRNA and each mRNA, we use the lower bound of all possible causal effect absolute values as the output.

For illustration, assume that we have the CPDAG G and a class of DAGs as in Fig. 2, where R_1 is a miRNA and T is a mRNA. We can use do-calculus to estimate the causal effect of R_1 on T in all four DAGs, G_1, G_2, G_3 and G_4. Suppose that the causal effects calculated based on the four DAGs are 0.5, 0.2, −0.3, and 0.4 respectively, then the value of 0.2 will be chosen as the final result. We use absolute values when comparing the strength of causal effects, as a calculated causal effect can be either negative or positive.

Details of the causal discovery procedure and its implementation in R can be found in [27].

2.1.3. Step (3): Identifying direct miRNA–mRNA causal regulatory relationships

In Step (2), we have found the causal effects for all possible pairs of miRNA–mRNA. Each obtained causal effect value indicates the strength of the regulation of a miRNA on a mRNA. However, the value does not tell whether the relationship between the miRNA and the mRNA is direct or mediated by other miRNAs or mRNAs.

A simple solution for differentiating direct from indirect relationships is to use the signs of causal effects obtained from Step (2). A positive causal effect means that when we manipulate the expression value of the miRNA to be increased, the expression value of a target gene will increase too. In other words, a positive causal effect implies up-regulation, and vice versa. As the common knowledge about miRNA regulation is that a miRNA usually down-regulates its target genes (only a few reported cases for up-regulation so far), we may simply assume that miRNA–mRNA relationships with negative causal effect values are direct interactions, and those with positive causal effects are indirect relationships. However, as human knowledge about miRNA regulation is still limited, the above assumption does not necessarily hold.

In this paper, we propose a solution that is based on target information predicted using sequence data. We argue that the binding site information about a target gene can provide evidence for direct miRNA–mRNA interactions. In other words, miRNA target information predicted from sequence data can be used as a means for differentiating direct and indirect miRNA–mRNA relationships. However, the predicted miRNA target information usually involves a high number of false discoveries. Therefore, to take advantage of and to compensate for the drawbacks of the target prediction and the causal discovery approaches, in Step (2) above, we use the causal discovery approach to find out the miRNA–mRNA causal relationships (with higher accuracy), and then in this step, from these discovered causal relationships, target information is used to select direct miRNA–mRNA interactions.

For the EMT case, we query Microcosm to retrieve target information for the differentially expressed miRNAs that were identified in Step (1). We use the target information to filter the miRNA–mRNA causal relationships discovered in Step (2). The miRNA–mRNA causal relationships that are not supported by the target information are removed. Then we rank all the remaining direct interactions for each miRNA based on the absolute values of its causal effects on mRNAs. The top interactions in the final result will have high causal effects and have been predicted using sequence data, and they are selected for validation (see the Results section).

2.2. Exploring TF–miRNA–mRNA causal feedforward patterns

To further explore the upstream regulators of miRNA regulation, we use the same steps described above to infer the direct TF–miRNA and TF–mRNA causal regulatory relationships, where TFs are considered as regulators while miRNAs and mRNAs are targets of the TFs. Then they are integrated with the direct miRNA–mRNA causal regulatory relationships found to form the causal feedfoward patterns (CFFPs). With each CFFP, a TF regulates a mRNA directly and indirectly via a miRNA.

To find all the TFs in the EMT datasets, we use the list of TF repertoire [38] to extract all the TF genes. This list is then used to query against the mRNA expression profiles for EMT to obtain TF expression profiles. The TF–miRNA target information is downloaded from MIR@NT@N [39]. To obtain TF–mRNA target information, we use TRANSFAC 9.3 [40] and promoter databases [41] integrated in the composite regulatory signature database (CRSD) [42].

Using DirectTarget, we infer all the direct causal regulatory relationships of TF–miRNA and TF–mRNA. Then we rank all the causal regulatory interactions of miRNA–mRNA (found previously),

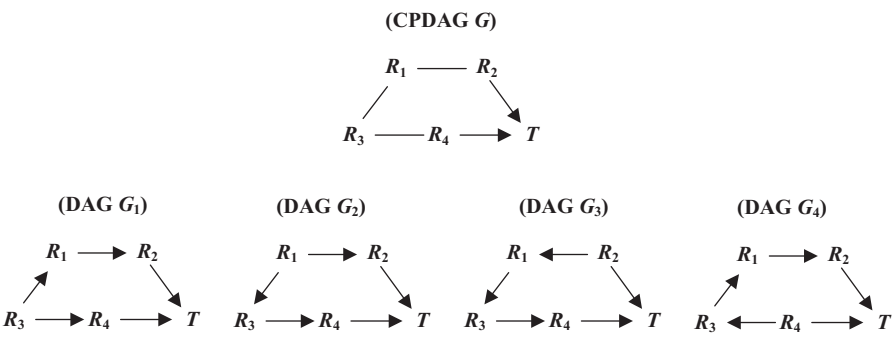

Fig. 2. A CPDAG G with four DAGs G_1, G_2, G_3 and G_4 in its *equivalence class* [28]. R_1, R_2, R_3 and R_4 denote four different miRNAs, and T represents a mRNA.

TF–miRNA, and TF–mRNA. Theoretically, any interactions with causal effects different from zero can be used to explore the TF–miRNA–mRNA causal feedforward patterns. In order to reduce false negatives (type II error) and for exploration purpose, we keep the interactions with 'above-the-middle' strength for the exploration. Specifically, we use the median of the absolute values of the causal effects as the cutoff to determine candidates of potential miRNA–mRNA, TF–miRNA, and TF–mRNA causal interactions. These causal interactions are then integrated to explore the causal feedforward patterns of which TFs are the upstream regulators. Details of the patterns are presented in the Results section.

3. Results

3.1. Data preparation and implementation

After the differentially expressed gene analysis of the EMT datasets using the *limma* package [35], 46 probes of miRNAs and 1612 probes of mRNAs are identified to be differentially expressed at a significant level (adjusted p-value < 0.05, adjusted by Benjamini–Hochberg (BH) method). We extracted 112 probes of TFs from the differentially expressed mRNAs using the list of TF repertoire [38]. The detailed result can be found in Supplementary material 1.

The input of DirectTarget is a 47×1658 matrix for the 47 NCI-60 samples of two different types. For inferring miRNA–mRNA causal regulatory interactions, the first 46 columns of the input data matrix are miRNA expression data, and the remaining 1612 columns are mRNA expression data. For exploring causal feedforward patterns of TF–miRNA–mRNA, the first 112 columns of the input data matrix are TF expression data, the next 46 columns are miRNA expression data, and the remaining 1500 columns are mRNA expression data.

The PC algorithm [36] is used to learn the causal structure from an input dataset. To estimate the completed partially directed acyclic graph (CPDAG), G from the input dataset, we use the open source R-package, *pcalg* [43], and set the significance level of the conditional independence test, $\alpha = 0.01$, as the tuning parameter for the PC algorithm. As described in the Methods section, for each causal regulatory relationship, the causal effect with respect to each of the DAGs in the equivalence class is calculated, and we regard the causal effect with the minimum absolute value as the final result.

Since a small number of samples can cause unstable estimations of causal effects, we use bootstrapping to estimate causal effects of the discovered causal regulatory relationships. The number of bootstrapping is set to 100 and the median of 100 estimates for each relationship is regarded as the final result.

3.2. Validated miRNA–mRNA interactions by experimentally confirmed target databases

To show the effectiveness of DirectTarget in identifying direct miRNA–mRNA interactions, we compare the performance of DirectTarget with the Pearson correlation method [24], Lasso [25], GenMiR++ [15], ProMISe [26] and the generic IDA method [27] respectively using the number of validated miRNA–mRNA interactions as the criterion. We define the ground truth for validating computational results as the union of the two largest experimentally confirmed target databases: miRTarBase v4.5 [44] and TarBase v6.0 [45]. Pearson correlation is computed using the R built-in function, *cor* (Package *stats*). We implement Lasso using *glmnet* [46] with the parameter *alpha* = 1, and remove those miRNA–mRNA interactions with Lasso's correlation coefficient of 0. The Matlab code of GenMiR++ is obtained from http://www.psi.toronto.edu/genmir/. To implement ProMISe, we use the average expression values across the 47 EMT samples of miRNAs and mRNAs as the input. Since the seed matrix between miRNAs and mRNAs is very sparse, which leads to failure in running ProMISe, we add a *ones* matrix based on the original seed matrix. As for generic IDA, we infer miRNA–mRNA causal regulatory relationships without using target binding information. To have a fair comparison of all methods, we extract from the discovered direct regulatory relationships the top 10 and 20 target genes of each validated miRNA respectively, based on the absolute values of the causal effects, correlations or probabilities. The detailed results of each method in identifying miRNA–mRNA regulatory relationships can be seen in Supplementary material 2.

Out of the 46 differentially expressed miRNAs in the EMT dataset, 44 are to be validated since 2 miRNAs do not have confirmed targets in the databases or do not have target binding information available. As shown in Fig. 3, DirectTarget is comparable with the Pearson correlation method, and performs better than Lasso, GenMiR++, ProMISe and generic IDA in terms of the total number of validated miRNA–mRNA interactions for the two cases (Top 10 and Top 20). This implies that DirectTarget can be used as a good alternative to existing methods for discovering target genes of miRNAs.

However, each method discovers a different set of confirmed miRNA–mRNA interactions, suggesting their own predicting merits. Fig. 4 shows the differences in the numbers of validated miRNA–mRNA interactions predicted by DirectTarget and the other five methods. In the top 10 lists for the 44 miRNAs, there are 6, 2, 6, 6 and 4 validated miRNA–mRNA interactions discovered by Pearson, Lasso, GeneMiR++, ProMISe, and generic IDA respectively, but missed by DirectTarget. However, there are 5, 12, 8, 10, and 17 validated miRNA–mRNA interactions identified by DirectTarget but missed by Pearson, Lasso, GeneMiR++, ProMISe, and generic IDA, respectively. In the top 20 lists of the 44 miRNAs, DirectTarget does not identify 4, 0, 11, 5 and 11 confirmed miRNA–mRNA interactions discovered by Pearson, Lasso, GeneMiR++, ProMISe, and generic IDA respectively. Nevertheless, in the top 20 lists, there are 6, 20, 12, 9 and 27 validated miRNA–mRNA interactions discovered by DirectTarget, but missed by Pearson, Lasso, GeneMiR++, ProMISe, and generic IDA respectively. The detailed information of validated miRNA–mRNA interactions by different methods is in Supplementary material 3.

In order to assess the statistical significance of the number of validated miRNA–mRNA interactions, we conceive a cumulative hypergeometric distribution model for this.

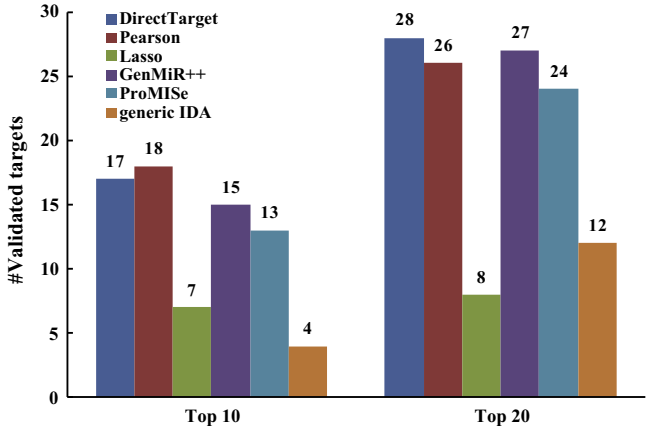

Fig. 3. A comparison of the total number of experimentally confirmed miRNA–mRNA interactions out of all the top 10 and top 20 target genes for the 44 miRNA used in the validation. The union of miRTarBase v4.5 and TarBase v6.0 is used as the ground truth. DirectTarget is comparable with the Pearson method, and outperforms the other four methods: Lasso, GenMiR++, ProMISe and generic IDA.

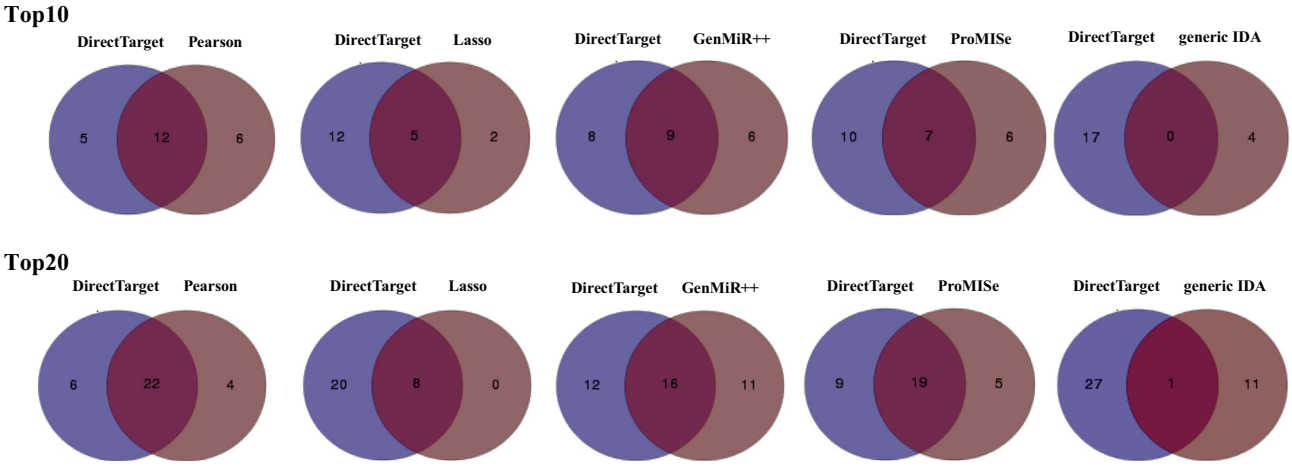

Fig. 4. The differences in the numbers of validated miRNA–mRNA interactions predicted by DirectTarget and the other five methods. Top10 and Top20 denote different sets of validated miRNA–mRNA interactions between DirectTarget and the other five methods in the top 10 and 20 lists for the 44 miRNAs, respectively.

Let S be the number of possible target genes for n miRNAs used in the validation in the EMT dataset, K be the number of miRNA–mRNA interactions from the ground truth for these miRNAs, N be the number of miRNA–mRNA interactions predicted by each method for these miRNAs, and x be the number of validated miRNA–mRNA interactions by the ground truth for these miRNAs. The p-value of the validation results, which is the probability of the random method equal to or better than each method, is calculated using the cumulative hypergeometric test formula:

$$p(X \geqslant x) = \sum_{i=x}^{N} \frac{\binom{K}{i}\binom{S*n-K}{N-i}}{\binom{S*n}{N}} \quad (1)$$

Table 1 shows the total number of validated miRNA–mRNA interactions out of all the top 10 and 20 lists of the 44 miRNAs obtained by each method together with the p-values for each of the results. The low p-values for DirectTarget in Table 1 suggest that the validation results of the method in discovering direct miRNA–mRNA causal regulatory relationships are statistically significant and not obtained by chance.

We also use the cumulative hypergeometric test to assess the statistical significance of the prediction for each single miRNA, i.e. using formula (1) when $n = 1$. With the assessment based on the top 10 list of each miRNA, the numbers of significant miRNAs, i.e. miRNAs with statistically significant predictions (p-value < 0.05) by the six methods (DirectTarget, Pearson, Lasso, GenMiR++, ProMISe and generic IDA) are 9, 8, 4, 6, 6 and 1 respectively. With the assessment based on the top 20 list of each miRNA, there are 7, 5, 1, 6, 5 and 1 significant miRNAs respectively for the six methods. The results show that DirectTarget could identify more miRNAs with validated targets that is statistically significant than the other five methods in both cases. The detailed information of the statistical significance of each miRNA by different methods is in Supplementary material 3.

3.3. Confirmed direct miRNA–mRNA causal interactions for EMT by literature

In order to balance the chance of finding more novel interactions and the cost of the exploration, instead of using the top-k interactions (e.g. $k = 100$), we use the median of the absolute values of all calculated causal effects as the cutoff in selecting potential causal relationships. When we list the interactions between miRNAs and mRNAs at probe level, a mRNA having different probe names may have multiple interactions with the same miRNA. Therefore when we represent the interactions at mRNA level, i.e. using gene symbols, there are duplicated miRNA–mRNA causal interactions. For each set of duplicated interactions, we keep only the interaction with the highest causal effect. As a result, we obtain 975 unique direct miRNA–mRNA causal regulatory relationships which involve 44 miRNAs and 493 unique mRNAs for EMT (details in Supplementary material 4).

Given that the miR-200 family is closely associated with epithelial to mesenchymal transition, we focus the exploration on direct causal regulatory relationships of the miR-200 family. As shown in Fig. 5 (red bold-faced line), the results indicate that ZEB1 is causally co-regulated by miR-200b, miR-200c and miR-429, and ZEB2 is co-regulated by miR-141, miR-200a, miR-200b, miR-200c and miR-429. Previous research [47–50] has shown that the miR-200 family inhibits the initiating step of metastasis and EMT by targeting the E-cadherin transcriptional repressors ZEB1 and ZEB2. The expression and inhibition of the miR-200 family causally

Table 1
Statistical significance of experimentally validated miRNA–mRNA interactions for the 44 miRNAs in the EMT dataset that are used in the validation. N_1 and N_2 are the top 10 and top 20 miRNA–mRNA interactions predicted by each method for the 44 miRNAs, respectively. Since those miRNA–mRNA interactions with Lasso's correlation coefficient of 0 are not used in the validation, the values of N_1 and N_2 for Lasso is 296 and 394 respectively. x_1 and x_2 denote the number of validated miRNA–mRNA interactions of each method in the two cases: Top 10 and Top 20, respectively. p_1 and p_2 represent p-values in the Top 10 and Top 20 cases, respectively.

Methods	S	n	K	$N_1(N_2)$	$x_1(x_2)$	$p_1(p_2)$
DirectTarget	1127	44	447	440(880)	17(28)	6.9965E−07(1.2481E−08)
Pearson	1127	44	447	440(880)	18(26)	1.4307E−07(1.7550E−07)
Lasso	1127	44	447	296(394)	7(8)	0.0186(0.0276)
GenMiR++	1127	44	447	440(880)	15(27)	1.3931E−05(4.7719E−08)
ProMISe	1127	44	447	440(880)	13(24)	2.1248E−04(2.1025E−06)
Generic IDA	1127	44	447	440(880)	4(12)	0.5616(0.1042)

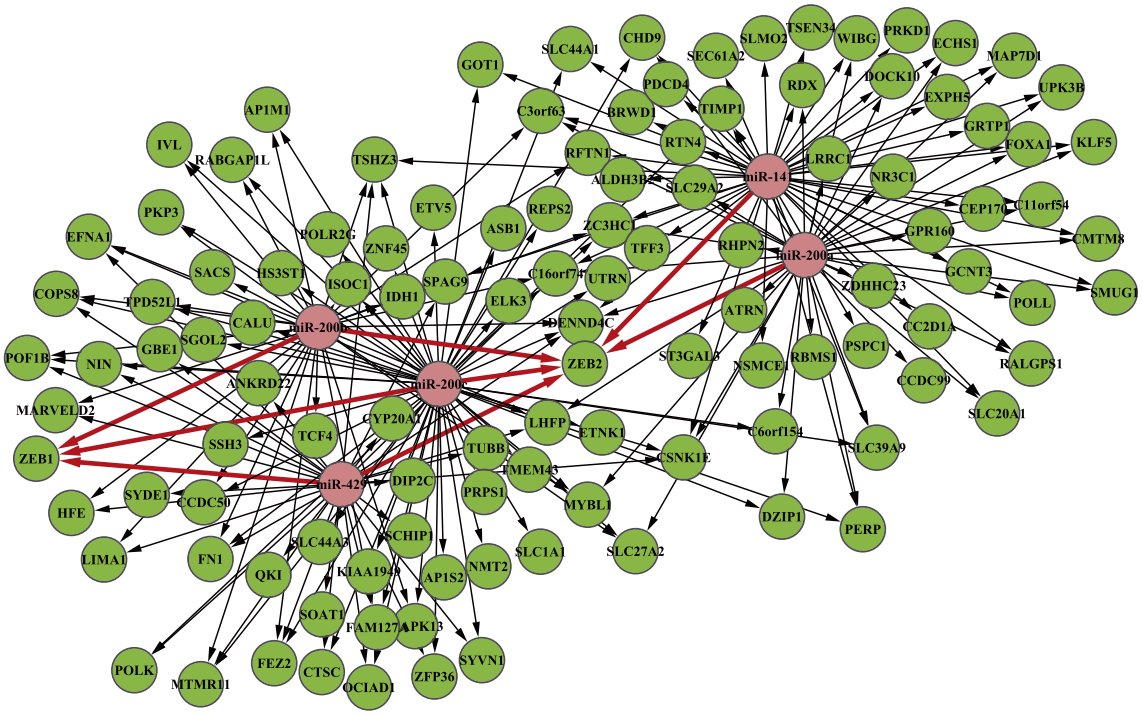

Fig. 5. Direct causal regulatory relationships of the miR-200 family (miR-141, miR-200a, miR-200b, miR-200c and miR-429). Red circles denote miRNAs and green circles are mRNAs. Red bold-faced lines are experimentally confirmed interactions. All of the nodes in the confirmed interactions are EMT bio-markers. They are the miR-200 family, ZEB1 and ZEB2. The miR-200 family that regulates ZEB1 and ZEB2 for EMT has been confirmed by literature. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

determine the regulation types of ZEB1 and ZEB2. The discoveries using DirectTarget are consistent with these results in the literature. Furthermore, ZEB1 and ZEB2 as bio-markers are also confirmed by our findings. The two mRNAs causally regulated by the miR-200 family are the markers in all three subtypes of EMT [51].

3.4. Functional validation of mRNAs for EMT

As the results are generated based on the EMT datasets, we extract the significant direct miRNA–mRNA causal regulatory relationships with the median of the causal effect (absolute) values as the cutoff, and validate those mRNAs as target genes for EMT. The functions of the target genes and the molecular pathways they potentially constitute are performed with the Ingenuity Pathway Analysis (IPA, Ingenuity Systems, www.ingenuity.com). Significant biological functions are identified for the target genes with a p-value cutoff of 0.05.

The causally regulated target genes are significantly enriched for several biological functions. The top five biological functions from IPA that are crucial for EMT are cancer, reproductive system disease, endocrine system disorders, cellular movement and dermatological diseases and conditions. Especially, the sub-categories of cellular movement, migration, invasion and scattering are critical for EMT since they have been identified as the functional markers of EMT in vitro [52]. As illustrated in Table 2 (details in Supplementary material 5), a significant number of target genes are associated with migration and invasion biological functions at significant level (max p-value < 0.05) and they are functional bio-markers for EMT.

3.5. Exploring causal feedforward patterns of TF–miRNA–mRNA

Great efforts have been made in investigating how miRNAs act in regulating target genes and their roles in various biological conditions. However, we have seen relatively few research into how miRNAs are regulated by TFs (TF–miRNA regulation). In fact, the genes with more TF-binding sites have a higher probability of being targeted by miRNAs and have more miRNA-binding sites on average, indicating that miRNAs and TFs may collaborate to regulate gene expression [53]. Therefore in this paper, we also examine the causal feedforward patterns (CFFPs) of TF–miRNA–mRNA in order to get insights into the indirect TF–mRNA regulation as a result of direct TF–miRNA and miRNA–mRNA causal interactions.

To explore the CFFPs, we only retain the direct TF–miRNA and miRNA–mRNA causal interactions with high absolute values of causal effects. We also use the median of the absolute values of the causal effects of all indirect TF–mRNA relationships as the cutoff to get significant CFFPs of TF–miRNA–mRNA. In total 1188 significant CFFPs of TF–miRNA–mRNA are selected, which involve 8 unique TFs, 29 miRNAs and 317 unique mRNAs for EMT (details in Supplementary material 6).

We focus on the significant CFFPs of TF–miRNA–mRNA that involve the miR-200 family (miR-141, miR-200a, miR-200b, miR-200c and miR-429) because some causal relationships in this category have been validated. As shown in Fig. 6, transcription factor ZEB2 can also directly regulate miR-200a, miR-200b and miR-429 (red bold-faced line). This implies that ZEB2 and the miR-200 family are reciprocally linked in a feedback loop, each one strictly controlling the other. This result is consistent with previous reports that a ZEB/miR-200 feedback loop plays a role in regulating EMT and cancer invasion [48,54]. Another transcriptional factor associated with EMT is SNAI2 (SLUG), and it plays an important role in EMT events during development, cancer and fibrosis [55]. Our results infer that SNAI2 directly regulates the miR-200 family transcript (red bold-faced line in Fig. 6). This discovery is consistent with the known discovery that SNAI2 regulates the miR-200 family [56].

We have also confirmed an observation that when the direct regulation of both TF–miRNA and miRNA–mRNA are up-regulation (positive causal effects) or down-regulation (negative causal

Table 2
Identified target genes that are significantly involved in Migration and Invasion biological functions and are functional bio-markers of EMT. The results are generated by IPA.

Functions	Molecules	#Molecules	Min *p*-value	Max *p*-value
Migration	ABL1, ACTR3, AGK, AKT3, ANPEP, AREG, ARHGAP8, CD9, CMTM8, CST6, CTSC, CTSL1, CTSL2, DAB2, DDR1, DLC1, EFNA1, EFNB3, ELK3, EPB41L5, EPHA1, EPHB3, ERBB2, ERBB3, ESRRA, ETV5, F2RL1, FLNA, FN1, FOLR1, FOXA1, GALNT2, GIT2, GNAI2, GNAS, GRB7, GRHL2, HAMP, IL17RB, ITGA5, ITGB4, JUP, KLF5, KLK3, KLK6, KRT6A, KRT8, LAMA3, LCN2, LIMA1, MCAM, MCF2L, MIEN1, MSN, MSX2, MYLK, MYLK3, NCOA3, NFIA, NINJ1, NMU, NOTCH4, PARD6B, PDCD4, PIK3C2B, PLXNB1, PRKCZ, PRKD1, PTPRM, RAP1GAP, REPS2, RLN2, RTN4, S100A14, S100B, S100P, SCHIP1, SCNN1A, SERPINB5, SGPP1, SLC37A4, SLC3A2, SLFN12L, SPAG9, SPDEF, SPG20, SREBF1, TACSTD2, TIMP1, TLN1, TNFRSF10A, TNFSF13, TNFSF13B, TNS4, TPT1, TRAF4, TUBA1A, TUBB2B, VIL1, VIM, ZEB1, ZEB2, ZFYVE21	103	2.41E−08	2.70E−03
Invasion	ABL1, AKT3, ANPEP, AREG, CD9, CLDN3, CMTM8, CST6, CTNND1, CTSL1, CTSL2, DAB2, DLC1, EFNA1, EFNB3, ELF3, ERBB2, ERBB3, ETV5, FLNA, FN1, GALNT2, GNAI2, GNAS, HIPK2, ITGA5, ITGB4, JUP, KLK3, KLK6, LCN2, MCAM, MIEN1, MSX2, MYLK, NCOA3, OTUB1, PDCD4, PLEKHG6, PLXNB1, RAP1GAP, S100A14, S100P, SERPINB5, SPDEF, ST14, TFF3, TIMP1, TNS4, VIM, ZEB1, ZEB2	52	5.56E−08	1.54E−03

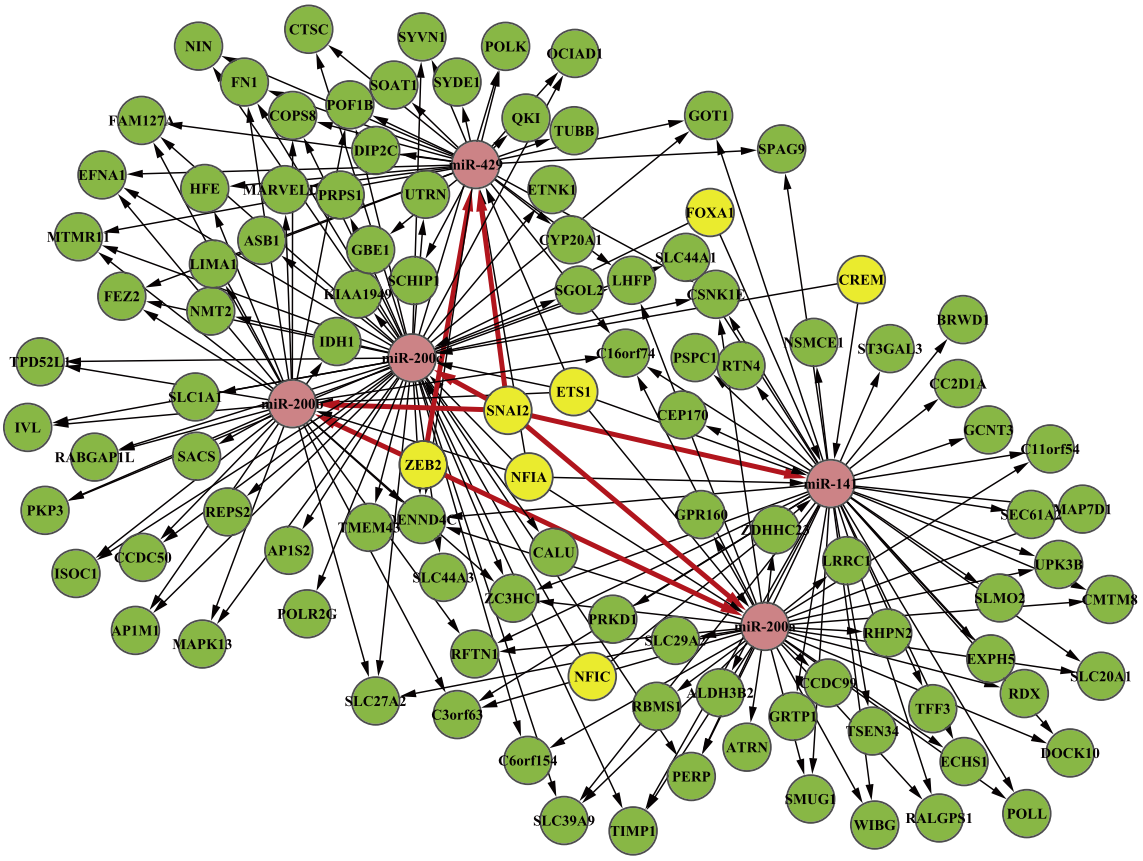

Fig. 6. Significant CFFPs of TF–miRNA–mRNA involving the miR-200 family (miR-141, miR-200a, miR-200b, miR-200c and miR-429). Yellow circles denote TFs, red circles denote miRNAs and green circles represent mRNAs. Red bold-faced lines are experimentally confirmed interactions. All of the nodes in the confirmed interactions are EMT bio-markers. ZEB2 and SNAI2 (SLUG) directly regulate the miR-200 family transcript, and ETS1 plays a crucial role in the process of EMT. These findings have been confirmed by literature. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

effects), the indirect regulation of TF–mRNA shows as up-regulation. Meanwhile, when the direct regulation of TF–miRNA is different from the regulation type of miRNA–mRNA, the indirect regulation of TF–mRNA shows as down-regulation. The consistency of the occurrences is up to 100% when the median of causal effect (absolute) values is used as the cutoff when selecting indirect TF–mRNA relationships for examination. Even when considering all indirect TF–mRNA causal relationships the consistency is up to 95.58% (details in Supplementary material 6). The results demonstrate that we can infer indirect regulation types of TF–mRNA using direct regulation types of TF–miRNA and miRNA–mRNA.

4. Conclusions and discussions

miRNAs are important gene regulators at the post-transcriptional level. However, it is still very challenging to achieve accurate prediction of the regulatory relationships between miRNAs and their target genes. In this paper, we have proposed a causal discovery based framework (DirectTarget) to identify the direct causal relationships between miRNAs and mRNAs. DirectTarget utilises both gene expression data and the target information predicted based on sequence data. The computational results of DirectTarget have been validated to be effective with experimentally confirmed target databases.

Most existing computational methods can only discover from observational data the statistical relationships between miRNAs and mRNAs, neglecting the causal nature of gene regulatory relationships. The recent work in [27] has adapted the idea of causal discovery for identifying causal regulatory relationships between miRNAs and mRNAs. However, the method is not able to distinguish direct causal regulatory relationships from indirect ones, while knowing whether a possible regulatory relationship is direct or indirect is essential for biologists when setting up lab experiments. On the other hand, miRNA target prediction based on sequence data is expected to provide the evidence of direct regulation of miRNAs on mRNA, but most target prediction methods based on sequence data have high false discovery rates and output too many possible targets, which makes the results less valuable in experiment design.

In this paper, we exploit the advantages of the causal discovery approach and sequence data based target prediction. DirectTarget firstly uses causal discovery to identify causal regulatory relationships between miRNAs and mRNAs with high accuracy, then we use predicted target information to guide the selection of direct regulatory relationships from the causal relationships identified in the first step. When applying DirectTarget to the EMT datasets, the validation results indicate that DirectTarget can effectively infer the direct causal relationships between miRNAs and their target genes. Moreover, the functional validations of the causally regulated target genes show that a significant number of target genes are highly associated with EMT. The experimental validations show that DirectTarget can be a useful tool to assist the experimental design for gene regulatory studies. Apart from the experimentally confirmed miRNA–mRNA relationships, the results generated by DirectTarget for other miRNAs still need further research and follow-up experiments.

Several methods [57–59] have been proposed to infer the regulatory relationships involving the three components, miRNAs, TFs, and mRNAs. However, to the best of our knowledge, there is no study specifically examining the direct causal regulatory relationships of TF–miRNA and miRNA–mRNA simultaneously and how these two types of direct regulatory relationships are connected. As a first attempt to this challenging problem, in this paper, we have also applied DirectTarget to explore the causal feedforward patterns of TF–miRNA–mRNA. The results show that TFs can indirectly regulate their target genes by directly regulating miRNAs.

DirectTarget provides a means to identify direct miRNA–mRNA causal regulatory relationships in heterogeneous data. The method is based on causality discovery and thus is different from existing correlation/association based approaches. As gene regulatory relationships are causal relationships, theoretically, DirectTarget is designed to return a correct set of miRNA–mRNA interactions when the number of samples is large. However, the large sample size requirement is often violated in real world datasets, and therefore the predictive power of DirectTarget is decreased.

The experimental results show that DirectTarget demonstrates some advantages over the other existing methods. Note that the validation results are based on currently available experimentally confirmed miRNA targets which are still far from complete. Moreover, each method discovered a unique set of validated interactions that other methods failed to identify, indicating that DirectTarget and other existing methods have their own merits in predicting miRNA targets.

In the future, we plan to tackle the small sample problem by integrating domain knowledge in causal structure learning. In this way, we may reduce the false identification of edges in a causal Bayesian network introduced by conditional independence tests with a small number of samples. The improved causal network will help achieve more accurate estimation of causal effects. We also plan to evaluate the performance of DirectTarget and other existing methods on larger cancer datasets from The Cancer Genome Atlas (TCGA, https://tcga-data.nci.nih.gov/tcga/tcgaHome2.jsp), where the number of samples for each cancer type is larger than what we used in this paper and the causality discovery based method will have more advantages than other methods.

Based on the results of this paper, nonetheless DirectTarget is a promising and complementary alternative to other existing methods for discovering target genes of miRNAs. In the future, we will further improve the computational efficiency of the method, and use more comprehensive miRNA target binding information to identify direct miRNA–mRNA causal regulatory relationships. Furthermore since many diseases are more likely caused by the effects of several miRNAs rather than a single miRNA [60], it is useful to infer the miRNAs synergistic effects and investigate gene regulation mechanisms at a system-wide level in the miRNA–miRNA synergistic networks. In the future we will construct miRNA–miRNA synergistic causal networks that reflect the co-regulation of miRNAs.

Authors contributions

JPZ, TDL and JYL conceived the idea of this work. LL and JH refined the idea. JPZ and TDL designed and performed the experiments. GG, BL and TDL provided the miRNA data and validated the results. JPZ, TDL, BL and JYL drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We are grateful to the reviewers for their constructive comments. This work has been partially supported by the Applied Basic Research Foundation of Science and Technology of Yunnan Province (No: 2013FD038), the Australian Research Council Discovery Grant DP130104090, and the Science Research Foundation for Youth Scholars of Dali University (No: KYQN201203).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jbi.2014.08.005.

References

[1] Kusenda B, Mraz M, Mayer J, Pospisilova S. MicroRNA biogenesis, functionality and cancer relevance. Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub 2006;150(2):205–15.
[2] Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009;136(2):215–33.
[3] Ambros V. MicroRNAs: tiny regulators with great potential. Cell 2001;107(7):823–6.
[4] Ambros V. MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell 2003;113(6):673–6.
[5] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004;116(2):281–97.
[6] Bushati N, Cohen SM. MicroRNA functions. Annu Rev Cell Dev Biol 2007;23:175–205.
[7] Du T, Zamore PD. Beginning to understand microRNA function. Cell Res 2007;17:661–3.
[8] Esquela-Kerscher A, Slack FJ. Oncomirs-microRNAs with a role in cancer. Nat Rev Cancer 2006;6:259–69.
[9] Cui Q, Yu Z, Purisima EO, Wang E. Principles of microRNA regulation of a human cellular signaling network. Mol Syst Biol 2006;2:46.
[10] Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG. The database of experimentally supported targets: a functional update of TarBase. Nucl Acids Res 2009;37:D155–8.
[11] Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ. miRBase: tools for microRNA genomics. Nucl Acids Res 2008;36:D154–8.
[12] Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell 2005;120(1):15–20.
[13] Krek A, Grün D, Poy MN, Wolf R, Rosenberg L, Epstein EJ, et al. Combinatorial microRNA target predictions. Nat Genet 2005;37(5):495–500.

[14] Rajewsky N. microRNA target predictions in animals. Nat Genet 2006;38(Suppl):S8–13.
[15] Huang JC, Babak T, Corson TW, Chua G, Khan S, Gallie BL, et al. Using expression profiling data to identify human microRNA targets. Nat Methods 2007;4(12):1045–9.
[16] Joung JG, Hwang KB, Nam JW, Kim SJ, Zhang BT. Discovery of microRNAmRNA modules via population-based probabilistic learning. Bioinformatics 2007;23(9):1141–7.
[17] Tran DH, Satou K, Ho TB. Finding microRNA regulatory modules in human genome using rule induction. BMC Bioinformatics 2008;9(suppl 12):S5.
[18] Joung JG, Fei Z. Computational identification of condition-specific miRNA targets based on gene expression profiles and sequence information. BMC Bioinformatics 2009;10(suppl 1):S34.
[19] Liu B, Li J, Tsykin A, Liu L, Gaur AB, Goodall GJ. Exploring complex miRNA–mRNA interactions with Bayesian networks by splitting–averaging strategy. BMC Bioinformatics 2009;10:408.
[20] Joung JG, Fei Z. Identification of microRNA regulatory modules in arabidopsis via a probabilistic graphical model. Bioinformatics 2009;25(3):387–93.
[21] Liu B, Li J, Tsykin A. Discovery of functional miRNA–mRNA regulatory modules with computational methods. J Biomed Inform 2009;42(4):685–91.
[22] Liu B, Liu L, Tsykin A, Goodall GJ, Green JE, Zhu M, et al. Identifying functional miRNA–mRNA regulatory modules with correspondence latent dirichlet allocation. Bioinformatics 2010;26(24):3105–11.
[23] Zhang J, Liu B, He J, Ma L, Li J. Inferring functional miRNA–mRNA regulatory modules in epithelial–mesenchymal transition with a probabilistic topic model. Comput Biol Med 2012;42(4):428–37.
[24] Fu J, Tang W, Du P, Wang G, Chen W, Li J, et al. Identifying microRNA–mRNA regulatory network in colorectal cancer by a combination of expression profile and bioinformatics analysis. BMC Syst Biol 2012;6:68.
[25] Lu Y, Zhou Y, Qu W, Deng M, Zhang C. A Lasso regression model for the construction of microRNA-target regulatory networks. Bioinformatics 2011;27(17):2406–13.
[26] Li Y, Liang C, Wong KC, Jin K, Zhang Z. Inferring probabilistic miRNA–mRNA interaction signatures in cancers: a role-switch approach. Nucl Acids Res 2014;42(9):e76.
[27] Le TD, Liu L, Tsykin A, Goodall GJ, Liu B, Sun BY, et al. Inferring microRNA–mRNA causal regulatory relationships from expression data. Bioinformatics 2013;29(6):765–71.
[28] Maathuis HM, Kalisch M, Buhlmann P. Estimating high-dimensional intervention effects from observational data. Ann Stat 2009;37(6):3133–64.
[29] Maathuis HM, Colombo D, Kalisch M, Buhlmann P. Predicting causal effects in large-scale systems from observational data. Nat Methods 2010;7(4):247–9.
[30] Savagner P. Leaving the neighborhood: molecular mechanisms involved during epithelial–mesenchymal transition. BioEssays 2001;23(10):912–23.
[31] Dvorak HF. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. New Engl J Med 1986;315(26):1650–9.
[32] Fuchs I, Lichtenegger W, Buehler H, Henrich W, Stein H, Kleine-Tebbe A, et al. The prognostic significance of epithelial–mesenchymal transition in breast cancer. Anticancer Res 2002;22(6A):3415.
[33] Park SM, Gaur AB, Lengyel E, Peter ME. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the Ecadherin repressors ZEB1 and ZEB2. Genes Dev 2008;22(7):894–907.
[34] Søkilde R, Kaczkowski B, Podolska A, Cirera S, Gorodkin J, Møller S, et al. Global microRNA analysis of the NCI-60 cancer cell panel. Mol Cancer Ther 2011;10(3):375–84.
[35] Smyth GK. Limma: linear models for microarray data. Bioinformatics and computational biology solutions using R and bioconductor. New York: Springer; 2005. pp. 397–420.
[36] Spirtes P, Glymour C, Scheines R. Causation, prediction, and search. 2nd ed. Cambridge, MA: MIT Press; 2000.
[37] Judea P. Causality: models, reasoning, and inference. Cambridge University Press; 2000.
[38] Vaquerizas JM, Kummerfeld SK, Teichmann SA, Luscombe NM. A census of human transcription factors: function, expression and evolution. Nat Rev Genet 2009;10(4):252–63.
[39] Le Béchec A, Portales-Casamar E, Vetter G, Moes M, Zindy PJ, Saumet A, et al. MIR@NT@N: a framework integrating transcription factors, microRNAs and their targets to identify sub-network motifs in a meta-regulation network model. BMC Bioinformatics 2011;12:67.
[40] Matys V, Fricke E, Geffers R, Gössling E, Haubrock M, Hehl R, et al. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucl Acids Res 2003;31:374–8.
[41] Halees AS, Weng Z. PromoSer: improvements to the algorithm, visualization and accessibility. Nucl Acids Res 2004;32(Web Server issue):W191–4.
[42] Liu CC, Lin CC, Chen WS, Chen HY, Chang PC, Chen JJ, et al. CRSD: a comprehensive web server for composite regulatory signature discovery. Nucl Acids Res 2006;34(Web Server issue):W571–7.
[43] Kalisch M, Mächler M, Colombo D, Maathuis MH, Bühlmann P. Causal inference using graphical models with the R package pcalg. J Stat Softw 2012;47(11):1–26.
[44] Hsu SD, Tseng YT, Shrestha S, Lin YL, Khaleel A, Chou CH, et al. miRTarBase update 2014: an information resource for experimentally validated miRNA-target interactions. Nucl Acids Res 2014;42(Database issue):D78–85.
[45] Vergoulis T, Vlachos IS, Alexiou P, Georgakilas G, Maragkakis M, Reczko M, et al. TarBase 6.0: capturing the exponential growth of miRNA targets with experimental support. Nucl Acids Res 2012;40(Database issue):D222–9.
[46] Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1–22.
[47] Gregory PA, Bracken CP, Bert AG, Goodall GJ. MicroRNAs as regulators of epithelial–mesenchymal transition. Cell Cycle 2008;7(20):3112–8.
[48] Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S, et al. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep 2008;9(6):582–9.
[49] Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol 2008;10:593–601.
[50] Korpal M, Lee ES, Hu G, Kang Y. The miR-200 family inhibits epithelial–mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. J Biol Chem 2008;283:14910–4.
[51] Zeisberg M, Neilson EG. Biomarkers for epithelial–mesenchymal transitions. J Clin Invest 2009;119(6):1429–37.
[52] Lee JM, Dedhar S, Kalluri R, Thompson EW. The epithelial–mesenchymal transition: new insights in signaling, development, and disease. J Cell Biol 2006;172(7):973–81.
[53] Cui Q, Yu Z, Pan Y, Purisima EO, Wang E. MicroRNAs preferentially target the genes with high transcriptional regulation complexity. Biochem Biophys Res Commun 2007;352(3):733–8.
[54] Bracken CP, Gregory PA, Kolesnikoff N, Bert AG, Wang J, Shannon MF, et al. A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial–mesenchymal transition. Cancer Res 2008;68(19):7846–54.
[55] Barrallo-Gimeno A, Nieto MA. The Snail genes as inducers of cell movement and survival: implications in development and cancer. Development (Cambridge, UK) 2005;132(14):3151–61.
[56] Liu YN, Yin JJ, Abou-Kheir W, Hynes PG, Casey OM, Fang L, et al. MiR-1 and miR-200 inhibit EMT via Slug-dependent and tumorigenesis via Slug-independent mechanisms. Oncogene 2013;32(3):296–306.
[57] Roqueiro D, Huang L, Dai Y. Identifying transcription factors and microRNAs as key regulators of pathways using Bayesian inference on known pathway structures. Proteome Sci 2012;10(suppl 1):S15.
[58] Zacher B, Abnaof K, Gade S, Younesi E, Tresch A, Frohlich H. Joint Bayesian inference of condition-specific miRNA and transcription factor activities from combined gene and microRNA expression data. Bioinformatics 2012;28(13):1714–20.
[59] Le TD, Liu L, Liu B, Tsykin A, Goodall GJ, Satou K, et al. Inferring microRNA and transcription factor regulatory networks in heterogeneous data. BMC Bioinformatics 2013;14:92.
[60] Xu J, Li CX, Li YS, Lv JY, Ma Y, Shao TT, et al. MiRNAmiRNA synergistic network: construction via co-regulating functional modules and disease miRNA topological features. Nucl Acids Res 2011;39(3):825–36.