# Identifying microRNA targets in epithelial-mesenchymal transition using joint-intervention causal inference

Thuc Duy Le*
School of Information Technology and Mathematical Sciences,
University of South Australia
Adelaide, Australia
Thuc.Le@unisa.edu.au

Junpeng Zhang†
School of Engineering, Dali University
Dali, Yunnan, China
zhangjunpeng_411@yahoo.com

Lin Liu
School of Information Technology and Mathematical Sciences,
University of South Australia
Adelaide, Australia
lin.liu@unisa.edu.au

Buu Minh Thanh Truong
Pham Ngoc Thach University of Medicine
Ho Chi Minh, Vietnam
trminhthanhbuu04@gmail.com

Shu Hu
State Key Laboratory of Information Security, Institute of Information Engineering, Chinese Academy of Sciences
Beijing, China
hushu@iie.ac.cn

Taosheng Xu
Institute of Intelligent Machines, Hefei Institutes of Physical Science, Chinese Academy of Sciences
Hefei, China
taosheng.x@gmail.com

Jiuyong Li‡
School of Information Technology and Mathematical Sciences,
University of South Australia
Adelaide, Australia
Jiuyong.Li@unisa.edu.au

## ABSTRACT

microRNAs (miRNAs) are important gene regulators, controlling a wide range of biological processes and being involved in several types of cancers. Currently, several computational approaches have been developed to elucidate the miRNA-mRNA regulatory relationships. However, these approaches have their own limitations and we are still far from understanding the miRNA-mRNA relationships, especially in specific biological processes. In this paper, we adapt a causal inference method to infer miRNA targets from the Epithelial Mesenchymal Transition (EMT) dataset. Our method utilises a causality based method that estimates the causal effect of each miRNA on a mRNA while controlling the effects of other miRNAs on the mRNA. The inferred causal effect is similar to the effect of a miRNA on a mRNA when we knockout all the other miRNAs. The experimental results show that our method is better than existing benchmark methods in finding experimentally confirmed miRNA targets. Moreover, we have found that the miR-200 family members (miR-141, miR-200a/b/c, and miR-429) synergistically regulate a number of target genes in EMT, suggesting their roles in controlling cancer metastasis. In addition, functional and pathway enrichment analyses show that the discovered miRNA-mRNA regulatory relationships are highly enriched in EMT, implying the validity of the proposed method. Novel miRNA-mRNA regulatory relationships discovered by our method provide a rich resource for follow up wet-lab experiments and EMT related studies.

*Thuc Duy Le co-first author and corresponding author.
†Junpeng Zhang co-first author.
‡To whom correspondence should be addressed.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
*CSBio '17, December 7–8, 2017, Nha Trang City, Viet Nam*
© 2017 Association for Computing Machinery.
ACM ISBN 978-1-4503-5350-2/17/12...$15.00
https://doi.org/10.1145/3156346.3156353

## CCS CONCEPTS

- **Applied computing → Bioinformatics**;

## KEYWORDS

miRNA, mRNA, Epithelial-mesenchymal transition, Causal inference

**ACM Reference Format:**
Thuc Duy Le, Junpeng Zhang, Lin Liu, Buu Minh Thanh Truong, Shu Hu, Taosheng Xu, and Jiuyong Li. 2017. Identifying microRNA targets in epithelial-mesenchymal transition using joint-intervention causal inference. In *Proceedings of 8th International Conference on Computational Systems-Biology and Bioinformatics, Nha Trang City, Viet Nam, December 7–8, 2017 (CSBio '17)*, 8 pages.
https://doi.org/10.1145/3156346.3156353

# 1 INTRODUCTION

Metastasis, a process whereby cells migrate away from the primary tumour, is the major cause of death from most cancers [10]. For example, breast cancer patients may die because of lung cancer metastasis. Cancer cells become metastatic and invasive through a process known as epithelial to mesenchymal transition (EMT) [32]. There has been evidence showing that EMT is controlled by microRNAs (miRNAs) [10], a class of short (∼22nt) endogenous non-coding RNAs that regulate gene expression by promoting messenger RNA (mRNA) degradation and repressing translation [3]. Therefore, inferring miRNA-mRNA interactions in EMT plays an important role in understanding cancer metastasis and assisting with the design of cancer treatments.

miRNAs recognise target mRNAs by base pairing to complementary sequences in the 3'-untranslated region (3'UTR), 5'UTR or sometimes in the open reading frames (ORFs) of the target mRNA [1, 3–5, 28]. miRNAs have been shown to regulate a wide range of biological processes, including proliferation [6, 48], metabolism [35], differentiation [9], development [11], apoptosis [43], cellular signaling [7] and even cancer development and progression [3, 16]. Recent findings have also shown that miRNAs not only interact with mRNAs, but with a pool of different RNAs, including long non-coding RNAs, circular RNAs, and pseudo genes in a complex network [23, 26]

There have been several methods developed to predict putative miRNA targets at the sequence level [8, 18, 25]. The common principles used in those methods are complementary base pairing, site accessibility, and evolutionary conservation. However, the results of those methods may contain a high-rate of false positives and false negatives [36].

With the advance in experimental technology, gene expression data has emerged as the important and promising resources for exploring miRNA functions. Various computational methods (see [22] for a review) have been devised to incorporate gene expression profiles into the study of miRNA-mRNA regulatory relationships. The principle of these methods is that if a gene is regulated by a miRNA, the dependency should show between the expression levels of the gene and the miRNA. These expression-based methods can be classifed as i) correlation-based approaches [27, 30, 41], and ii) causal inference approaches [21, 46, 47].

All of these approaches possess their own limitations. The correlation-based approaches are based-on the anti-correlation property between the expression levels of the miRNAs and the mRNAs. However, correlations or associations are not causality, whereas the miRNA-mRNA regulatory relationships are causal. For instance, a strong anti-correlation between the expression values of a miRNA and a mRNA in a dataset may be confounded by another factor, e.g. a transcription factor, and thus it is a spurious relationship. In another direction, causal inference approaches estimate the causal effects of miRNAs on the mRNAs, i.e. simulating the intervention effects in the gene knockdown experiments. While the causal inference approaches, e.g. [21, 46, 47], help remove spurious relationships, the estimated causal effects only show the total effect that one miRNA has on a mRNA, i.e. the effect of miRNA $X$ on mRNA $Y$ may contain not only the contribution of miRNA X, but also the contributions of other miRNAs (e.g. those along the paths from miRNA X to mRNA $Y$). Inferring the pure causal effect of each miRNA on a mRNA is important for the understanding of the roles of each miRNA in EMT and cancer progression. Such causal effects can be achieved in a lab by transfecting (manipulating) a miRNA of interest, while knocking-down all other miRNAs at the same time. However, it is impractical to conduct the experiments for all miRNAs of interest, given there are hundreds of miRNAs need to be tested.

In this paper, we propose to infer the pure causal effects of miRNAs on mRNAs by utilising the joint-intervention causal effect estimation approach [31]. Our goal is to identify the causal effect of a miRNA on a mRNA while adjusting for the effects coming from other miRNAs. Generally speaking, we simulate the experiment of manipulating one miRNA of interest and knocking-down all other miRNAs at the same time, and we aim to estimate the effect of the miRNA of interest on a mRNA. We have applied the method to the NCI-60 panel cell lines dataset to infer the miRNA-mRNA relationships in EMT. Experimental results show that our proposed method finds more experimentally confirmed miRNA targets than the benchmark methods, including the existing causal inference method [21]. In addition, functional enrichment analyses show that the identified miRNA-mRNA relationships are highly enriched in functions and processes related to EMT, suggesting the usefulness of the method. Novel interactions identified by the proposed method could be good candidates for follow-up wet-lab experiments to explore their roles in EMT.

# 2 METHODS

## 2.1 Notations

We use the same graph notations as in [21, 40].

Let $G = (\mathbf{V}, \mathbf{E})$ be a graph consisting of a set of vertices $\mathbf{V}$ and a set of edges $\mathbf{E} \subseteq \mathbf{V} \times \mathbf{V}$. In our context, $\mathbf{V} = \{X_1, \ldots, X_p\}$ is a set of random variables representing the expression levels of miRNAs and mRNAs, and the edges represent the regulatory relationships between these variables. Graph $G$ is a Directed Acyclic Graph (DAG) if $G$ contains only directed edges and has no cycles. The *skeleton* of a DAG $G$ is the undirected graph obtained from $G$ by substituting undirected edges for directed edges. A *v-structure* is an ordered triple of vertices, $(X_i, X_j, X_k)$, such that in $G$ there exist directed edges $X_i \to X_j$ and $X_j \leftarrow X_k$, and $X_i$ and $X_k$ are not adjacent. $X_j$ is then known as a *collider* in this $v$-structure. An *equivalence class* of DAGs is the set of DAGs which have the same skeleton and the same $v$-structures.

An equivalence class of DAGs can be uniquely described by a *completed partially directed acyclic graph* (CPDAG). A partially directed acyclic graph (PDAG) is a graph where the edges are either directed or undirected and one cannot trace a cycle by following the directions of the directed edges and any directions of the undirected edges. A PDAG is *completed* if

(1) every directed edge exists also in every DAG belonging to the equivalence class; (2) for every undirected edge, $X_i - X_k$, there exists a DAG with $X_i \leftarrow X_k$ and a DAG with $X_i \rightarrow X_k$ in the equivalence class.

## 2.2 Estimating multiple-intervention effects

The aim of estimating joint-intervention effect of $X_1, ..., X_m$ on $Y$ is to estimate $(\beta_1, ..., \beta_m)$, where $\beta_i, i \in \{1, ..., m\}$ is the causal effect of $X_i$ on $Y$ when we knockdown all $X_j, j \neq i$ at the same time.

The causal effects $(\beta_1, ..., \beta_m)$ here reflect the amount of change in the target $Y$ when we change (manipulate) one unit of $X_i, i \in \{1, ..., m\}$. Following Pearl's *do-calculus* [33], the joint causal effect of $X_1, ..., X_m$ on $Y$ is defined as the vector:

$(\beta_1, ..., \beta_m) = (E[Y|do(X_1 = x_1 + 1, X_2 = x_2, ..., X_m = x_m)] - E[Y|do(X_1 = x_1, ..., X_m = x_m)], ..., E[Y|do(X_1 = x_1, ..., X_m = x_m + 1)] - E[Y|do(X_1 = x_1, ..., X_m = x_m)])$,

where $do(X_i = x_i)$ is the operation to force $X_i$ to receive the value $x_i$.

If we denote the miRNAs as $(X_1, ..., X_m)$ and the mRNAs as $(Y_1, ..., Y_n)$, we will be able to estimate the joint effect $(\beta_1, ..., \beta_k)$ of $(X_1, ..., X_m)$ on each of $Y_j, j \in \{1, ..., n\}$ using the above method. However, in this work, our aim is to identify the target mRNAs of each miRNA (excluding the effects of other miRNAs) rather than finding the total joint effects of multiple miRNAs. To this end, we only utilise the value of each individual $\beta_i, i \in \{1, ..., m\}$, which represents the causal effect of miRNA $X_i$ on the mRNA $Y_j$ when we manipulating all $\{X_j\}_{j \neq i}$, to rank the target mRNAs of a miRNA. Therefore, the vector $(\beta_1, ..., \beta_m)$ provides us sufficient information for estimating the effect of each miRNA on each target mRNA.

However, *do-calculus* requires a causal structure of the variables to be given as a DAG, which is often unknown in reality. To bridge the gap, Nandy et al [31] proposed a method to estimate the joint causal effects from observational data alone. The method includes two main phases: (i) to learn a causal structure from observational data and (ii) to apply the recursive regressions for causal effects (RRC) method [31] to infer the joint causal effects. The method is implemented in *R* and available in the *pcalg* package [13].

## 2.3 Identifying miRNA targets using joint-intervention effects

Based on the above discussion, we propose an algorithm for identifying miRNA targets, which comprises the following 4 steps.

**Step 1**: Identify significant miRNAs and mRNAs. We use differently expressed analysis to find significantly differentially expressed miRNAs and mRNAs between Epithelial and Mesenchymal samples.

**Step 2**: Use the PC algorithm [40] to learn the CPDAG $G$ whose nodes represent the differently expressed miRNAs and mRNAs. We use the parallelised version of the algorithm [19] to reduce the running time.

**Step 3**: Estimate the causal effects of each miRNA on each mRNA by controlling the effects of other miRNAs. For each mRNA, we use the joint intevention effect estimation method in Nandy et al. [31] to estimate the $\beta_i, i \in \{1, ..., m\}$, which is the pure causal effect of each of the $m$ differently expressed miRNAs on each of the mRNAs.

**Step 4**: Refine the above estimated miRNA causal effects. The causal effects in Step 3 may have multiple values, as we estimated the effects from different DAGs in the PCDAG. In this step, we select the causal effect value with the smallest absolute value, and output it as the causal effect of the miRNA on the mRNA.

## 2.4 Implementation

The major step of the proposed method is to estimate the joint intervention effects, and this step is implemented in the R package *pcalg* ([12], *joinIda* function). However, it is not efficient to apply the method to gene expression datasets with thousands of variables. We implemented the parallelised jointIDA algorithm which uses multiple core CPU to speed up the runtime of the jointIDA algorithm. The software is available in our R package *ParallelPC* [20] in CRAN. The R script of jointIDA for reproducing the results in this paper is also available upon request.

## 2.5 Functional annotation of miRNAs

We use functional enrichment analysis of target genes to annotate the biological functions of miRNAs. The Disease Ontology (DO) [15] enrichment analysis is performed using the R package *DOSE* [45]. The Gene Ontology (GO) [2] and Kyoto Encyclopedia of Genes and Genomes (KEGG) [14] enrichment analysis are conducted using the R package *clusterProfiler* [44]. The DO, GO, KEGG terms with adjusted $p$-value $< 0.05$ (adjusted by Benjamini-Hochberg (BH) method) are regarded as functional categories.

# 3 RESULTS

## 3.1 Data preparation

The matched miRNA and mRNA expression profiles of the NCI-60 cancer cell panel are obtained from Søkilde *et al.* [39] and Shankavaram *et al.* [37], respectively. According to the work in [32], we classify the samples into epithelial (11 samples) and mesenchymal (36 samples).

In this work, we are interested in identifying regulatory relationships between differentially expressed miRNAs and mRNAs, and the *limma* package [38] of Bioconductor is applied for differential expression analysis. As a result, we obtain 35 probes of miRNAs and 1154 probes of mRNAs which are found to be differentially expressed at significant level (adjusted $p$-value $< 0.05$, adjusted by BH method). The matched expression data of the differentially expressed miRNAs and mRNAs can be found in Supplementary file 1 at nugget.unisa.edu.au/Thuc/CSBIO2017.

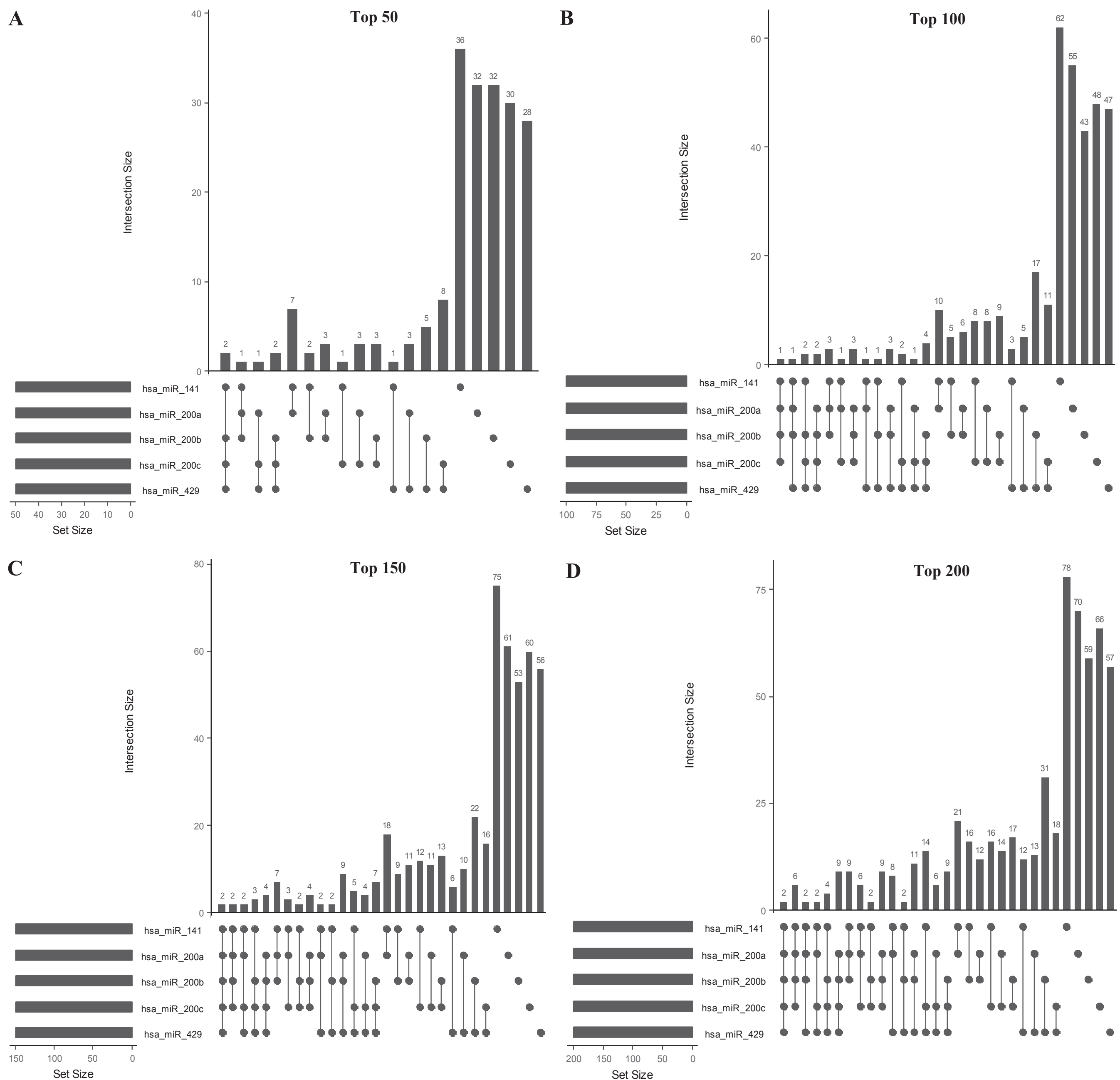


**Figure 1: Overlap and difference of top miRNA-miRNA interactions for each miRNA. (A) Top 50 miRNA-miRNA interactions for each miRNA. (B) Top 100 miRNA-miRNA interactions for each miRNA. (C) Top 150 miRNA-miRNA interactions for each miRNA. (D) Top 200 miRNA-miRNA interactions for each miRNA.**

To validate the predicted miRNA-mRNA interactions, we use the experimentally confirmed database TarBase [42] as the ground-truth for validation. Moreover, we also use the transfection data from [23] for validation. The transfection data can be obtained at nugget.unisa.edu.au/Thuc/miRLAB/logFC.imputed.rda.

## 3.2 The miR-200 family members synergistically regulate a quantity of target genes

To evaluate the strength of each predicted miRNA-mRNA regulation, we use the absolute value of causal effects which miRNAs have on mRNAs (Step 4 in the algorithm). The larger

**Figure 2: Visulisation of top 50 miRNA-mRNA interactions for the miR-200 family members. The red nodes and white nodes are miRNAs and mRNAs, respectively. The red lines denote experimentally validated miRNA-mRNA interactions.**

absolute value of causal effect, the stronger miRNA-mRNA regulation. To identify more reliable predicted miRNA-mRNA regulations, we empirically select top $k$ (50, 100, 150 and 200) miRNA-mRNA interactions for each miRNA. The top 50, 100, 150 and 200 miRNA-mRNA interactions for each miRNA can be seen in Supplementary file 2 at nugget.unisa.edu.au/Thuc/CSBIO2017.

Previous studies [10, 17, 29] have demonstrated that the miR-200 family members (miR-141, miR-200a/b/c, and miR-429) are closely associated with EMT by direct regulating

**Table 1: Several miRNAs are significantly enriched in cell migration.**

| Terms | Functions associated with cell migration | GO terms | miRNAs |
|---|---|---|---|
| Top 50 | epithelial cell migration | GO:0010631 | hsa-miR-1180, hsa-miR-33b |
| | negative regulation of epithelial cell migration | GO:0010633 | hsa-miR-1180 |
| | positive regulation of cell migration | GO:0030335 | hsa-miR-18b, hsa-miR-33b |
| | regulation of epithelial cell migration | GO:0010632 | hsa-miR-33b |
| Top 100 | ameboidal-type cell migration | GO:0001667 | hsa-miR-203, hsa-miR-32, hsa-miR-33b, hsa-miR-7-1* |
| | blood vessel endothelial cell migration | GO:0043534 | hsa-miR-200a* |
| | endothelial cell migration | GO:0043542 | hsa-miR-200a*, hsa-miR-30e |
| | epithelial cell migration | GO:0010631 | hsa-miR-200a*, hsa-miR-200b, hsa-miR-30e, hsa-miR-33b |
| | positive regulation of cell migration | GO:0030335 | hsa-miR-7-1* |
| | regulation of epithelial cell migration | GO:0010632 | hsa-miR-301b, hsa-miR-33b |
| | substrate-dependent cell migration | GO:0006929 | hsa-miR-200a* |

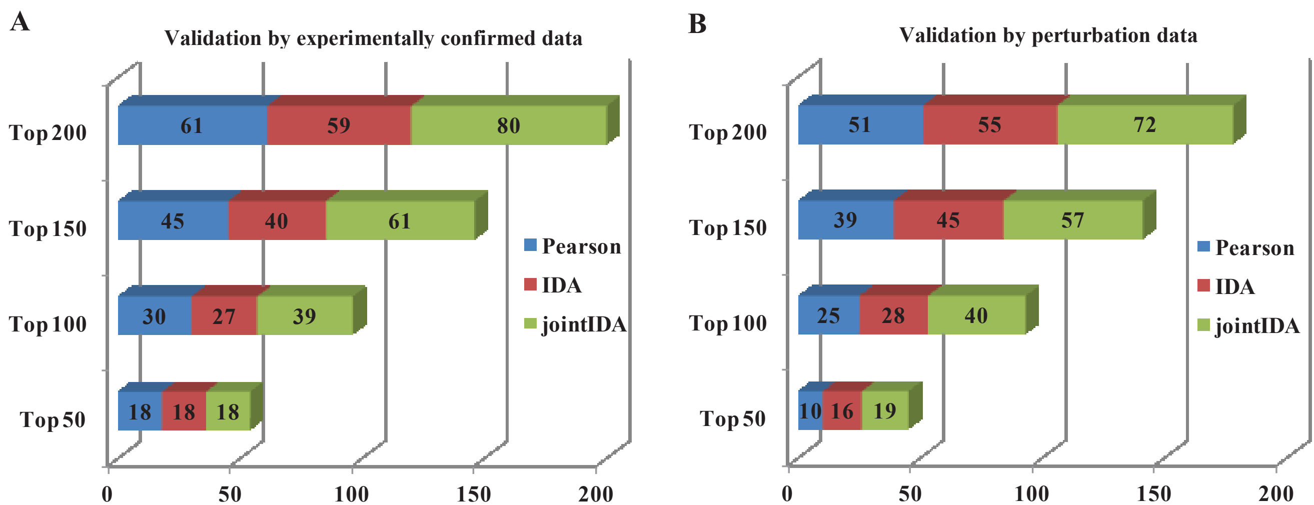


**Figure 3: Comparison of jointIDA, IDA and Pearson in predicting validated miRNA targets. The top 50, 100, 150, 200 miRNA-mRNA interactions for each miRNA are considered. The experimentally confirmed data (A) and perturbation data (B) are used for validation.**

target genes. Therefore, in this section, we focus on investigating miRNA-mRNA regulations associated with the miR-200 family members. As shown in Figure 1, a number of target genes are shared by at least two members of the miR-200 family in the four cases (top 50, 100, 150 and 200 target genes for each miRNA). This result suggests that the miR-200 family members tend to synergistically regulate a quantity of target genes. For a case study, in the top 50 miRNA-mRNA interactions for the miR-200 family members, 42 target genes are co-regulated by the miR-200 family members (Figure 1A). Figure 2 shows the visulization of top 50 miRNA-mRNA interactions for the miR-200 family members, and the miR-200 family members are connected by their co-regulated target genes. Among three validated target genes (CDH1, FN1 and ZEB2) of the miR-200 family members, two target genes (FN1 and ZEB2) are synergistically regulated.

## 3.3 Functional enrichment analysis of miRNA target genes

To uncover potential diseases and biological processes associated with miRNAs, we conduct functional enrichment analysis of miRNA target genes. Due to no mapped names of 6 miRNAs (hsa-miRPlus-A1018, hsa-miRPlus-A1055, hsa-miRPlus-A1065, hsa-miRPlus-B1030, hsa-miRPlus-D1054 and hsa-miRPlus-E1281) in miRBase, we remove the enrichment analysis results of them. In total, we perform functional enrichment analysis of target genes of 29 miRNAs. Moreover, we only select top 50 and 100 miRNA-mRNA interactions for each miRNA to make functional enrichment analysis, because the enrichment analysis results of hundrends of target genes are too general to gain biological insight. As a result, we discover that 23 out of the 29 miRNAs (79.31%) and 26 out of the 29 miRNAs (89.66%) in

terms of the top 50 and 100 miRNA-mRNA interactions respectively, are significantly associated with at least one DO, GO or KEGG term (the number of enriched DO, GO and KEGG terms for each miRNA can be seen in Supplementary file 3 at nugget.unisa.edu.au/Thuc/CSBIO2017). This result indicates that most miRNAs are biological meaningful in the EMT dataset.

Functional enrichment analysis results show that miRNAs are significantly enriched in several biological functions, including three sub-categories of cellular movement (cell migration, cell invasion, and cell scattering), which are critical for EMT [24]. As shown in Table 1, in the top 50 miRNA-mRNA interactions for each miRNA, hsa-miR-1180, hsa-miR-18b and hsa-miR-33b are functionally associated with cell migration. Moreover, in the top 100 miRNA-mRNA interactions for each miRNA, 8 miRNAs (hsa-miR-200a*, hsa-miR-200b, hsa-miR-203, hsa-miR-301b, hsa-miR-30e, hsa-miR-32, hsa-miR-33b, and hsa-miR-7-1*) are also found to be significantly associated with cell migration. The details of enrichment analysis results can be found in Supplementary file 4 at nugget.unisa.edu.au/Thuc/CSBIO2017.

### 3.4 Comparison results

In this section, we compare our method with the Pearson [34] and IDA [21] methods in terms of validated miRNA-mRNA interactions. For each method, we choose the top 50, 100, 150, 200 miRNA-mRNA interactions for each miRNA for validation. As shown in Figure 3, using both the experimentally confirmed data and perturbation data, the proposed method (jointIDA) always performs better than the other methods in predicting validated miRNA-mRNA interactions. The result demonstrates that jointIDA is a useful method for identifying miRNA targets.

## 4 CONCLUSIONS

miRNAs play important roles in plants, animals and some viruses by interacting with target genes and controlling key biological processes. In this paper, we aim to explore the miRNA-mRNA regulatory relationships in EMT using a causal inference method.

Although there have been serveral methods in identifying miRNA targets from expression data, existing methods mainly find the correlations or associations between the expression levels of miRNAs and mRNAs, which can not reflect the cause-effect nature of miRNA-mRNA regulations. Thus, it is important to identify miRNA-mRNA causal relationships by simulating gene knockdown experiments. In this paper, we propose a causal inference method, based on joint intervention effects [31], to identify miRNA-mRNA regualtory relationships. Unlike the existing causal method in [21], which used the single intervention effects, the proposed method estimates the causal effect of each miRNA on each mRNA while controlling the effects of other miRNAs. We apply the method to the EMT dataset, and the results show that the proposed method outperforms benchmark correlation-based and causallity-based methods in identifying miRNA targets. Functional pathway enrichment analyses show that the discovered results are highly relevant to EMT, suggesting the validity and usefulness of the method.

In the future, we will explore the applications of the proposed method in finding miRNA synergy networks and investigating the miRNA roles in cancer prognosis.

## ACKNOWLEDGMENTS

This work is supported by the National Health and Medical Research Council under Grant No.: 1123042; the Australian Research Council Discovery under Grant No.: DP170101306; the National Science Foundation of China under Grant No.: 61702069; and the Applied Basic Research Foundation of Science and Technology of Yunnan Province under Grant No.: 2017FB099.

## REFERENCES


[1] Victor Ambros. 2004. The functions of animal microRNAs. *Nature* (2004), 350–355.

[2] M Ashburner, C A Ball, J A Blake, D Botstein, H Butler, J M Cherry, A P Davis, K Dolinski, S S Dwight, and J T Eppig. 2000. Gene ontology: Tool for the unification of biology. The gene ontology consortium. *Nature Genetics* 25 (2000), 25–29. https://doi.org/10.1038/75556

[3] D P Bartel. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116 (2004), 281–297. https://doi.org/10.1016/S0092-8674(04)00045-5

[4] David P Bartel. 2009. MicroRNAs : Target recognition and regulatory functions. *Cell* 136 (2009), 215–233. https://doi.org/10.1016/j.cell.2009.01.002

[5] Eugene Berezikov, Edwin Cuppen, and Ronald H A Plasterk. 2006. Approaches to microRNA discovery. *Nature Genetics* 38, May (2006), 2–8. https://doi.org/10.1038/ng1794

[6] Jian-Fu Chen, Elizabeth M Mandel, J Michael Thomson, Qiulian Wu, Thomas E Callis, Scott M Hammond, Frank L Conlon, and Da-Zhi Wang. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nature Genetics* 38, 2 (Feb. 2006), 228–33. https://doi.org/10.1038/ng1725

[7] Qinghua Cui, Zhenbao Yu, Enrico O Purisima, and Edwin Wang. 2006. Principles of microRNA regulation of a human cellular signaling network. *Molecular Systems Biology* 2 (Jan. 2006), 1–7. https://doi.org/10.1038/msb4100089

[8] Anton J Enright, Bino John, Ulrike Gaul, Thomas Tuschl, Chris Sander, Debora S Marks, et al. 2004. microRNA targets in Drosophila. *Genome Biology* 5, 1 (2004), R1–R1.

[9] Aurora Esquela-Kerscher and Frank J Slack. 2006. Oncomirs - microRNAs with a role in cancer. *Nature Reviews. Cancer* 6, 4 (April 2006), 259–69. https://doi.org/10.1038/nrc1840

[10] Philip a Gregory, Andrew G Bert, Emily L Paterson, Simon C Barry, Anna Tsykin, Gelareh Farshid, Mathew a Vadas, Yeesim Khew-Goodall, and Gregory J Goodall. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biology* 10, 5 (May 2008), 593–601. https://doi.org/10.1038/ncb1722

[11] Peng Jin, Daniela C Zarnescu, Stephanie Ceman, Mika Nakamoto, Julie Mowrey, Thomas a Jongens, David L Nelson, Kevin Moses, and Stephen T Warren. 2004. Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nature Neuroscience* 7, 2 (Feb. 2004), 113–7. https://doi.org/10.1038/nn1174

[12] Markus Kalisch, Martin Machler, and Diego Colombo. 2010. pcalg: estimation of CPDAG/PAG and causal inference using the IDA algorithm. (2010).

[13] Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H. Maathuis, and Peter Bühlmann. 2012. Causal Inference Using Graphical Models with the R Package pcalg. *Journal of Statistical Software* 47, 11 (2012), 1–26. http://www.jstatsoft.org/v47/i11/

[14] Minoru Kanehisa and Susumu Goto. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* 28, 1 (2000), 27–30.

[15] Warren A Kibbe, Cesar Arze, Victor Felix, Elvira Mitraka, Evan Bolton, Gang Fu, Christopher J Mungall, Janos X Binder, James Malone, Drashtti Vasant, et al. 2014. Disease Ontology 2015 update: an expanded and updated database of human diseases for linking biomedical knowledge through disease data. *Nucleic acids research* 43, D1 (2014), D1071–D1078.
[16] V N Kim and J W Nam. 2006. Genomics of microRNA. *Trends Genetics* 22 (2006), 165–173. https://doi.org/10.1016/j.tig.2006.01.003
[17] Manav Korpal, Esther S Lee, Guohong Hu, and Yibin Kang. 2008. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *The Journal of Biological Chemistry* 283, 22 (May 2008), 14910–4. https://doi.org/10.1074/jbc.C800074200
[18] 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. 2005. Combinatorial microRNA target predictions. *Nature Genetics* 37, 5 (2005), 495–500.
[19] Thuc Le, Tao Hoang, Jiuyong Li, Lin Liu, Huawen Liu, and Shu Hu. 2016. A fast PC algorithm for high dimensional causal discovery with multi-core PCs. *IEEE/ACM transactions on computational biology and bioinformatics* (2016).
[20] Thuc Duy Le, Tao Hoang, Jiuyong Li, Lin Liu, and Shu Hu. 2015. ParallelPC: an R package for efficient constraint based causal exploration. *arXiv preprint arXiv:1510.03042* (2015).
[21] Thuc Duy Le, Lin Liu, Anna Tsykin, Gregory J Goodall, Bing Liu, Bing-Yu Sun, and Jiuyong Li. 2013. Inferring microRNA–mRNA causal regulatory relationships from expression data. *Bioinformatics* 29, 6 (2013), 765–771.
[22] Thuc Duy Le, Lin Liu, Junpeng Zhang, Bing Liu, and Jiuyong Li. 2014. From miRNA regulation to miRNA–TF co-regulation: computational approaches and challenges. *Briefings in bioinformatics* 16, 3 (2014), 475–496.
[23] Thuc Duy Le, Junpeng Zhang, Lin Liu, Huawen Liu, and Jiuyong Li. 2016. miRLAB: An R Based Dry Lab for Exploring miRNA-mRNA Regulatory Relationships. *PLoS One* 10, 12 (1 2016), e0145386. https://doi.org/10.1371/journal.pone.0145386
[24] Jonathan M Lee, Shoukat Dedhar, Raghu Kalluri, and Erik W Thompson. 2006. The epithelial-mesenchymal transition: new insights in signaling, development, and disease. *The Journal of Cell Biology* 172, 7 (March 2006), 973–81. https://doi.org/10.1083/jcb.200601018
[25] B P Lewis, C B Burge, and D P Bartel. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120 (2005), 15–20. https://doi.org/10.1016/j.cell.2004.12.035
[26] Jun-Hao Li, Shun Liu, Hui Zhou, Liang-Hu Qu, and Jian-Hua Yang. 2014. starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein–RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research* 42, D1 (2014), D92–D97.
[27] Huiqing Liu, Angela Brannon, Anupama Reddy, Gabriela Alexe, Michael Seiler, Alexandra Arreola, Jay Oza, Ming Yao, David Juan, Louis Liou, et al. 2010. Identifying mRNA targets of microRNA dysregulated in cancer: with application to clear cell Renal Cell Carcinoma. *BMC Systems Biology* 4, 1 (2010), 51.
[28] Gunter Meister and Thomas Tuschl. 2004. Mechanisms of gene silencing by double-stranded RNA. *Nature* 431, 7006 (Sept. 2004), 343–9. https://doi.org/10.1038/nature02873
[29] Perry S Mongroo and Anil K Rustgi. 2010. The role of the miR-200 family in epithelial-mesenchymal transition. *Cancer biology & therapy* 10, 3 (2010), 219–222.
[30] Ander Muniategui, Jon Pey, Francisco J Planes, and Angel Rubio. 2012. Joint analysis of miRNA and mRNA expression data. *Briefings in bioinformatics* 14, 3 (2012), 263–278.
[31] Preetam Nandy, Marloes H Maathuis, Thomas S Richardson, et al. 2017. Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *The Annals of Statistics* 45, 2 (2017), 647–674.
[32] Sun-Mi Park, Arti B Gaur, Ernst Lengyel, and Marcus E Peter. 2008. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes & Development* 22, 7 (April 2008), 894–907. https://doi.org/10.1101/gad.1640608
[33] Judea Pearl. 2000. *Causality: models, reasoning, and inference.* Cambridge University Press.
[34] Karl Pearson. 1920. Notes on the history of correlation. *Biometrika* 13, 1 (1920), 25–45.
[35] Matthew N Poy, Lena Eliasson, Jan Krutzfeldt, Satoru Kuwajima, Xiaosong Ma, Patrick E Macdonald, Sébastien Pfeffer, Thomas Tuschl, Nikolaus Rajewsky, Patrik Rorsman, and Markus Stoffel. 2004. A pancreatic islet-specific microRNA regulates insulin secretion. *Nature* 432, 7014 (Nov. 2004), 226–30. https://doi.org/10.1038/nature03076
[36] N Rajewsky. 2006. microRNA target predictions in animals. *Nature Genetics* 38, Suppl (2006), S8–13.
[37] Uma T Shankavaram, William C Reinhold, Satoshi Nishizuka, Sylvia Major, Daisaku Morita, Krishna K Chary, Mark A Reimers, Uwe Scherf, Ari Kahn, Douglas Dolginow, Jeffrey Cossman, Eric P Kaldjian, Dominic A Scudiero, Emanuel Petricoin, Lance Liotta, Jae K Lee, and John N Weinstein. 2007. Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Molecular cancer therapeutics* 6, 3 (March 2007), 820âĂŞ832. https://doi.org/10.1158/1535-7163.mct-06-0650
[38] Gordon K Smyth. 2005. Limma : Linear models for microarray data. *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (2005), 397–420.
[39] Rolf Søkilde, Bogumil Kaczkowski, and Agnieszka Podolska. 2011. Global microRNA analysis of the NCI-60 cancer cell panel. *Molecular Cancer Therapeutics* 10 (2011), 375–384. https://doi.org/10.1158/1535-7163.MCT-10-0605
[40] Peter Spirtes, C Glymour, and R Scheines. 2000. *Causation, prediction, and search* (2nd ed.). MIT Press, Cambridge, MA.
[41] Dang Hung Tran, Kenji Satou, and Tu Bao Ho. 2008. Finding microRNA regulatory modules in human genome using rule induction. *BMC Bioinformatics* 9 Suppl 12 (Jan. 2008), S5. https://doi.org/10.1186/1471-2105-9-S12-S5
[42] Ioannis S. Vlachos, Maria D. Paraskevopoulou, Dimitra Karagkouni, Georgios Georgakilas, Thanasis Vergoulis, Ilias Kanellos, Ioannis-Laertis Anastasopoulos, Sofia Maniou, Konstantina Karathanou, Despina Kalfakakou, Athanasios Fevgas, Theodore Dalamagas, and Artemis G. Hatzigeorgiou. 2015. DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions. *Nucleic Acids Research* 43, D1 (2015), D153–D159. https://doi.org/10.1093/nar/gku1215
[43] Chaoqian Xu, Yanjie Lu, Zhenwei Pan, Wenfeng Chu, Xiaobin Luo, Huixian Lin, Jiening Xiao, Hongli Shan, Zhiguo Wang, and Baofeng Yang. 2007. The muscle-specific microRNAs miR-1 and miR-133 produce opposing effects on apoptosis by targeting HSP60, HSP70 and caspase-9 in cardiomyocytes. *Journal of Cell Science* 120, Pt 17 (Sept. 2007), 3045–52. https://doi.org/10.1242/jcs.010728
[44] Guangchuang Yu, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics: a journal of integrative biology* 16, 5 (2012), 284–287.
[45] Guangchuang Yu, Li-Gen Wang, Guang-Rong Yan, and Qing-Yu He. 2014. DOSE: an R/Bioconductor package for disease ontology semantic and enrichment analysis. *Bioinformatics* 31, 4 (2014), 608–609.
[46] Junpeng Zhang, Thuc Duy Le, Lin Liu, Bing Liu, Jianfeng He, Gregory J Goodall, and Jiuyong Li. 2014. Identifying direct miRNA–mRNA causal regulatory relationships in heterogeneous data. *Journal of biomedical informatics* 52 (2014), 438–447.
[47] Junpeng Zhang, Thuc Duy Le, Lin Liu, Bing Liu, Jianfeng He, Gregory J Goodall, and Jiuyong Li. 2014. Inferring condition-specific miRNA activity from matched miRNA and mRNA expression data. *Bioinformatics* 30, 21 (2014), 3070–3077.
[48] Yong Zhao, Eva Samal, and Deepak Srivastava. 2005. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436, 7048 (July 2005), 214–20. https://doi.org/10.1038/nature03817